IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Art Unit 2188

JUN 1 8 2004 F

In re U.S. Patent Application of

KASAKO

Application Number: 10/765,128

Filed: January 28, 2004

DATA PROCESSING SYSTEM HAVING A

PLURALITY OF STORAGE SYSTEMS

Atty Docket No.

WILL.0001

Honorable Assistant Commissioner for Patents Washington, D.C. 20231

REQUEST FOR PRIORITY UNDER 35 U.S.C. § 119 AND THE INTERNATIONAL CONVENTION

Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of December 15, 2003, the filing date of the corresponding Japanese Patent Application 2003-416414.

A certified copy of Japanese Patent Application 2003-416414 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

Stanley P. Fisher

Registration Number 24,344

huan Carlos A Marquez

Registration Number 34,072

REED SMITH LLP

3110 Fairview Park Drive **Suite 1400** Falls Church, Virginia 22042 (703) 641-4200 June 18, 2004



JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application:

2003年12月15日

願 出 Application Number:

特願2003-416414

[ST. 10/C]:

Applicant(s):

[| P 2 0 0 3 - 4 1 6 4 1 4]

出 願 人

株式会社日立製作所

2004年 1月27日

特許庁長官 Commissioner, Japan Patent Office



【書類名】 特許願 【整理番号】 340301444 【あて先】 特許庁長官殿 【国際特許分類】 G06F 12/16 G06F 03/06

【発明者】

【住所又は居所】 神奈川県小田原市中里322番2号 株式会社日立製作所RAI

Dシステム事業部内

【氏名】 加迫 尚久

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100095371

【弁理士】

【氏名又は名称】 上村 輝之

【選任した代理人】

【識別番号】 100089277

【弁理士】

【氏名又は名称】 宮川 長夫

【選任した代理人】

【識別番号】 100104891

【弁理士】

【氏名又は名称】 中村 猛

【手数料の表示】

【予納台帳番号】 043557 【納付金額】 21,000円

【提出物件の目録】

【物件名】 特許請求の範囲 1

 【物件名】
 明細書 1

 【物件名】
 図面 1

 【物件名】
 要約書 1

 【包括委任状番号】
 0110323

【書類名】特許請求の範囲

【請求項1】

上位装置に通信可能に接続された第一の記憶システムと、

前記第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備え、

(1) 前記第一の記憶システムは、

前記上位装置から送られたデータが格納される第一のデータ記憶領域と、

前記第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域と、

前記上位装置から送られたデータを前記第一のデータ記憶領域に書き、前記第一のデータ記憶領域に書かれたデータのジャーナルを前記第一のジャーナル記憶領域に書き、前記第二及び第三の記憶システムのそれぞれからの要求に応じて前記第一のジャーナル記憶領域内の前記ジャーナルを前記第二及び第三の記憶システムにそれぞれ送るようになった第一の制御装置と

を有しており、

(2) 前記第二の記憶システムは、

前記第一のデータ記憶領域内のデータの複製が格納される第二のデータ記憶領域と、前記ジャーナルが格納される第二のジャーナル記憶領域と、

独自にスケジュールされたジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第二のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで前記第二のジャーナル記憶領域内の前記ジャーナルに基づいて前記第一のデータ記憶領域内のデータの複製を生成して前記第二のデータ記憶領域に書くようになった第二の制御装置とを有し、

(3) 前記第三の記憶システムは、

前記第一のデータ記憶領域内のデータの複製が格納される第三のデータ記憶領域と、 前記ジャーナルが格納される第三のジャーナル記憶領域と、

独自にスケジュールされたジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第三のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで前記第三のジャーナル記憶領域内の前記ジャーナルに基づいて前記第一のデータ記憶領域内のデータの複製を生成して前記第三のデータ記憶領域に書くようになった第三の制御装置とを有し、

前記第一の記憶システムの前記第一の制御装置は、前記第一のジャーナル記憶領域内の前記ジャーナルが前記第二及び第三の記憶システムによって読まれたか否かを検出し、前記第二及び第三の記憶システムの双方によって読まれるまでは前記第一のジャーナル記憶領域内の前記ジャーナルを保持し、前記第二及び第三の記憶システムの双方によって読まれた後に前記第一のジャーナル記憶領域内の前記ジャーナルを消去可能とするようになっている、

ことを特徴とするデータ処理システム。

【請求項2】

請求項1に記載のデータ処理システムにおいて、

(1) 前記第一の記憶システムは、複数の物理的記憶装置を有し、

前記第一の記憶システムの前記第一の制御装置は、前記上位装置との間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、

前記第一の制御装置は、前記第一の記憶システム内の前記複数の物理的記憶装置が持つ記憶領域を、前記第一のデータ記憶領域及び前記第一のジャーナル記憶領域に割当て、

(2) 前記第二の記憶システムは、複数の物理的記憶装置を有し、

2/

前記第二の記憶システムの前記第二の制御装置は、前記第一の記憶システムとの間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、

前記第二の制御装置は、前記第二の記憶システム内の前記複数の物理的記憶装置が持つ 記憶領域を、前記第二のデータ記憶領域及び前記第二のジャーナル記憶領域に割当て、

(2) 前記第三の記憶システムは、複数の物理的記憶装置を有し、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムとの間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、

前記第三の制御装置は、前記第三の記憶システム内の前記複数の物理的記憶装置が持つ記憶領域を、前記第三のデータ記憶領域及び前記第三のジャーナル記憶領域に割当てるようになっている、

ことを特徴とするデータ処理システム。

【請求項3】

請求項1に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一のジャーナル記憶領域から 読んだジャーナルのデータの数に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項4】

請求項1に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項5】

請求項1記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三のデータ記憶領域が保持している前記ジャーナルの記憶容量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項6】

請求項1に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三の記憶システムの処理負荷に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項7】

請求項1に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システム内の前記第一のジャーナル記憶領域が保持している前記ジャーナルの記憶容量についての情報を、前記第一の記憶システムから読み出し、前記読み出された前記ジャーナルの記憶容量についての情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項8】

請求項1に記載のデータ処理システムにおいて、

前記第一の記憶システムは、前記第一のジャーナル記憶領域についての管理情報を所有し、

前記第三の記憶システムの前記三の制御装置は、前記第一の記憶システムが所有している前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムから読み出し、読み出された前記第一のジャーナル記憶領域についての管理情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項9】

請求項1に記載のデータ処理システムにおいて、

前記第一の記憶システム内の前記第一のデータ記憶領域は、複数の論理ボリュームを有し、

前記第一の制御装置は、前記複数の論理ボリュームに格納される複数のデータにそれぞれ対応する複数のジャーナルを、前記第一のジャーナル記憶領域に書き、

前記第一のジャーナル記憶領域に格納される前記複数のジャーナルには、前記複数のジャーナルがそれぞれ対応する前記複数のデータの更新順序に関する情報が含まれており、

前記第二及び第三の記憶システムの前記第二及び第三の制御装置は、それぞれ、前記第一の記憶システムから読んだ前記複数のジャーナルに含まれている前記更新順序に従って、前記複数のジャーナルに基づいて前記複数のデータの複製を生成して前記第二及び第三のデータ記憶領域にそれぞれ書くことを特徴とするデータ処理システム。

【請求項10】

請求項1に記載のデータ処理システムにおいて、

前記第三の記憶システムは、前記第三の記憶システムの処理負荷に応じて、前記リストアのタイミングを制御することを特徴とするデータ処理システム。

【請求項11】

上位装置に通信可能に接続された第一の記憶システムと、

前記第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備え、

(1) 前記第一の記憶システムは、

前記上位装置から送られたデータが格納される第一のデータ記憶領域と、

前記第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域と、

前記上位装置から送られたデータを前記第一のデータ記憶領域に書き、そして、書かれたデータのジャーナルを前記第一のジャーナル記憶領域に書くようになった第一の制御装置とを有し、

(2) 前記第二の記憶システムは、

前記ジャーナルが格納される第二のジャーナル記憶領域と、

所定のジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを 読み、読まれたジャーナルを前記第二のジャーナル記憶領域に書くようになった第二の制 御装置とを有し、

(3) 前記第三の記憶システムは、

前記ジャーナルが格納される第三のジャーナル記憶領域と、

所定のジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを 読み、読まれたジャーナルを前記第三のジャーナル記憶領域に書くようになった第三の制 御装置とを有する、

ことを特徴とするデータ処理システム。

【請求項12】

請求項11記載のデータ処理システムにおいて、

前記第二の記憶システムは、前記データの複製を格納する第二のデータ記憶領域を更に有し、前記第二の制御装置が、所定のリストアのタイミングで前記第二のジャーナル記憶領域に格納されたジャーナルから前記データの複製を生成し、生成された前記データの複製を前記第二のデータ記憶装置に書くようになっており、

前記第三の記憶システムは、前記データの複製を格納する第三のデータ記憶領域を更に有し、前記第三の制御装置が、所定のリストアのタイミングで前記第三のジャーナル記憶領域に格納されたジャーナルから前記データの複製を生成し、生成された前記データの複製を前記第三のデータ記憶装置に書くようになっている、

ことを特徴とするデータ処理システム。

【請求項13】

請求項11記載のデータ処理システムにおいて、

前記第一の記憶システムの前記第一の制御装置は、前記第一のジャーナル記憶領域内の前記ジャーナルが前記第二及び第三の記憶システムによって読まれたか否かを検出し、前記第二及び第三の記憶システムの双方によって読まれるまでは前記第一のジャーナル記憶領域内の前記ジャーナルを保持し、前記第二及び第三の記憶システムの双方によって読まれた後に前記第一のジャーナル記憶領域内の前記ジャーナルを消去可能とするようになっている、

ことを特徴とするデータ処理システム。

【請求項14】

請求項11に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一のジャーナル記憶領域から 読んだジャーナルのデータの数に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項15】

請求項11に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項16】

請求項11記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三のデータ記憶領域が保持している前記ジャーナルの記憶容量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項17】

請求項11に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三の記憶システムの処理負荷に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項18】

請求項11に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システム内の前記第一のジャーナル記憶領域が保持している前記ジャーナルの記憶容量についての情報を、前記第一の記憶システムから読み出し、前記読み出された前記ジャーナルの記憶容量についての情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項19】

請求項11に記載のデータ処理システムにおいて、

前記第一の記憶システムは、前記第一のジャーナル記憶領域についての管理情報を所有 し、

前記第三の記憶システムの前記三の制御装置は、前記第一の記憶システムが所有している前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムから読み出し、読み出された前記第一のジャーナル記憶領域についての管理情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【書類名】明細書

【発明の名称】複数の記憶システムを有するデータ処理システム

【技術分野】

$[0\ 0\ 0\ 1]$

本発明は、複数の記憶システムを有するデータ処理システムに関し、特に複数の記憶システム間でのデータの複製に関する。

【背景技術】

[0002]

近年、常に顧客に対して継続したサービスを提供するために、第一の記憶システムに障害が発生した場合でもデータ処理システムがサービスを提供できるよう、記憶システム間でのデータの複製に関する技術が重要になっている。第一の記憶システムに格納された情報を第二および第三の記憶システムに複製する技術として、以下の特許文献に開示された技術が存在する。

[0003]

米国特許5170480号公報には、第一の記憶システムに接続された第一の計算機が、第一の記憶システムに格納されたデータを、第一の計算機と第二の計算機間の通信リンクを介し、第二の計算機に転送し、第二の計算機が第二の計算機に接続された第二の記憶システムに転送する技術が開示されている。

$[0\ 0\ 0\ 4]$

米国特許6209002号公報には、第一の記憶システムが、第一の記憶システムに格納されたデータを、第二の記憶システムに転送し、さらに、第二の記憶システムが、第三の記憶システムに転送する技術が開示されている。計算機と第一の記憶システムとは通信リンクにより接続され、第一の記憶システムと第二の記憶システムとは通信リンクにより接続され、さらに、第二の記憶システムと第三の記憶システムとは通信リンクにより接続されている。第一の記憶システムは、複製対象の第一の論理ボリュームを保持する。第二の記憶システムは、第一の論理ボリュームの複製である第二の論理ボリューム、及び、第二の論理ボリュームの複製である第四の論理ボリュームを保持する。この特許文献において、第二の記憶システムは、第二の論理ボリュームを保持する。この特許文献において、第二の記憶システムは、第二の論理ボリュームから第三の論理ボリュームへのデータ複製処理と、第三の論理ボリュームから第四の論理ボリュームへのデータ複製処理と、第三の論理ボリュームから第四の論理ボリュームへのデータ複製処理とを排他的に実行する。

[0005]

【特許文献1】米国特許5170480号公報

【特許文献2】米国特許6209002号公報

【発明の開示】

【発明が解決しようとする課題】

[0006]

米国特許 5 1 7 0 4 8 0 号公報に開示された技術は、データの複製のために、第一の計算機および第二の計算機を常に使用する。第一の計算機は、通常業務を行っており、第一の計算機にかかるデータ複製処理の負荷は無視できない。さらに、複製のためのデータは、第一の計算機と第一の記憶システム間の通信リンクを使用するため、通常業務のために必要なデータ転送と衝突し、通常業務に必要なデータ参照、データ更新時間が長くなるという課題がある。

[0007]

米国特許6209002号公報に開示された技術は、複製を行うデータ量の倍の記憶容量が第二の記憶システム及び第三の記憶システムに必要となる。また、複製対象のデータ量が多いことにより、データ複製処理に費やされる時間が長く、第三の記憶システムのデータは古いものとなってしまう。その結果、第三の記憶システムのデータを用いて業務が再開される場合、第三の記憶システムのデータを最新とするまでの時間が長くなり、業務再開までの時間が延びるという課題がある。さらに、この文献の開示によれば、第一の記

憶システムは、第一の記憶システム内のデータ更新処理に加えて、第二の記憶システムとの間でのデータ更新処理が終了した時点で、上位の計算機にデータ更新完了報告を行う。したがって、計算機からのデータ更新に費やされる時間が長く、第一の記憶システムと第二の記憶システムとの間の距離が遠くなればなるほど、データ更新に費やされる時間はより長くなる。その結果、各記憶システム間の距離をあまり遠くすることができないという問題もある。

[0008]

本発明の目的は、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製を行えるようにすることにある。

[0009]

本発明の別の目的は、記憶システムと計算機との間の通信に影響を与えずに、複数の記憶システム間でデータ転送又はデータの複製をすることにある。

$[0\ 0\ 1\ 0]$

本発明のまた別の目的は、複数の記憶システム内に保持されるデータ格納領域を少なく することにある。

$[0\ 0\ 1\ 1]$

本発明のさらに別の目的は、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることにある。

【課題を解決するための手段】

$[0\ 0\ 1\ 2]$

本発明の一つの観点に従うデータ処理システムは、上位装置に通信可能に接続された第一の記憶システムと、第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備える。

$[0\ 0\ 1\ 3\]$

第一の記憶システムは、第一の制御装置と、上位装置から送られたデータが格納される 第一のデータ記憶領域と、第一の記憶領域に格納されたデータの複製を生成するために使 用されるジャーナルが格納される第一のジャーナル記憶領域とを有する。第一の制御装置 は、上位装置から送られたデータを第一のデータ記憶領域に書き、第一のデータ記憶領域 に書かれたデータのジャーナルを第一のジャーナル記憶領域に書き、そして、第二及び第 三の記憶システムのそれぞれからの要求に応じて、第一のジャーナル記憶領域内のジャー ナルを第二及び第三の記憶システムにそれぞれ送るようになっている。

$[0\ 0\ 1\ 4\]$

第二の記憶システムは、第二の制御装置と、第一のデータ記憶領域内のデータの複製が格納される第二のデータ記憶領域と、ジャーナルが格納される第二のジャーナル記憶領域とを有する。第二の制御装置は、独自にスケジュールされたジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第二のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで、第二のジャーナル記憶領域内の前記ジャーナルに基づいて第一のデータ記憶領域内のデータの複製を生成して、これを第二のデータ記憶領域に書くようになっている。

$[0\ 0\ 1\ 5]$

三の記憶システムは、第三の制御装置と、第一のデータ記憶領域内のデータの複製が格納される第三のデータ記憶領域と、ジャーナルが格納される第三のジャーナル記憶領域とを有する。第三の制御装置は、独自にスケジュールされたジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第三のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで第三のジャーナル記憶領域内のジャーナルに基づいて第一のデータ記憶領域内のデータの複製を生成して、これを第三のデータ記憶領域に書くようになっている。

$[0\ 0\ 1\ 6]$

さらに、第一の記憶システム内の第一の制御装置は、第一のジャーナル記憶領域内のジ

ャーナルが第二及び第三の記憶システムによって読まれたか否かを検出する。そして、第一の制御装置は、第二及び第三の記憶システムの双方によって読まれるまでは、第一のジャーナル記憶領域内のジャーナルを保持し、第二及び第三の記憶システムの双方によって読まれた後に、第一のジャーナル記憶領域内のジャーナルを消去可能とするようになっている。

$[0\ 0\ 1\ 7]$

第三の記憶システムの第三の制御装置は、第一のジャーナル記憶領域から読んだジャー ナルのデータの数に応じて、ジャーナルリードの時間間隔を制御するようになっていてよ い。或いは、第三の制御装置は、第一の記憶システムと第三の記憶システムとの間で送受 されるデータの通信量に応じて、ジャーナルリードの時間間隔を制御するようになってい てもよい。或いは、第三の制御装置は、第三のデータ記憶領域が保持しているジャーナル の記憶容量に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。 或いは、第三の制御装置は、第三の記憶システムの処理負荷に応じて、ジャーナルリード の時間間隔を制御するようになっていてもよい。或いは、第三の制御装置は、第一の記憶 システム内の第一のジャーナル記憶領域が保持しているジャーナルの記憶容量についての 情報を、第一の記憶システムから読み出し、読み出されたジャーナルの記憶容量について の情報に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。或い は、第一の記憶システムが、第一のジャーナル記憶領域についての管理情報を所有してお り、そして、第三の記憶システムの第三の制御装置が、前記第一の記憶システムが所有し ている前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムか ら読み出し、読み出された第一のジャーナル記憶領域についての管理情報に応じて、ジャ ーナルリードの時間間隔を制御するようにしてもよい。第二の記憶システムの第二の制御 装置も、基本的に、これと同様にジャーナルリードの時間間隔を制御するようになってい てよい。

[0018]

第一の記憶システム内の第一のデータ記憶領域は、複数の論理ボリュームから構成されることができる。第一の制御装置は、その複数の論理ボリュームに格納される複数のデータにそれぞれ対応する複数のジャーナルを、第一のジャーナル記憶領域に書くことができる。第一のジャーナル記憶領域に格納される複数のジャーナルには、対応する複数のデータの更新順序に関する情報が含まれることができる。第二及び第三の記憶システムの第二及び第三の制御装置は、それぞれ、第一の記憶システムから読んだ複数のジャーナルに含まれている更新順序に従って、複数のジャーナルに基づいて複数のデータの複製を生成して、それを第二及び第三のデータ記憶領域にそれぞれ書くことができる。

$[0\ 0\ 1\ 9]$

第三の記憶システムは、第三の記憶システムの処理負荷に応じて、リストアのタイミングを制御するようにしてもよい。

[0020]

本発明の別の観点に従う上位装置に通信可能に接続された第一の記憶システムと、第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備える。

[0021]

第一の記憶システムは、第一の制御装置と、上位装置から送られたデータが格納される 第一のデータ記憶領域と、第一の記憶領域に格納されたデータの複製を生成するために使 用されるジャーナルが格納される第一のジャーナル記憶領域とを有する。この第一の記憶 システムの第一の制御装置は、上位装置から送られたデータを第一のデータ記憶領域に書 き、そして、書かれたデータのジャーナルを第一のジャーナル記憶領域に書くようになっ ている。

[0022]

第二の記憶システムは、第二の制御装置と、ジャーナルが格納される第二のジャーナル 記憶領域とを有する。この第二の記憶システムの第二の制御装置は、所定のジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第 二のジャーナル記憶領域に書くようになっている。

[0023]

第三の記憶システムは、第三の制御装置と、ジャーナルが格納される第三のジャーナル記憶領域とを有する。この第三の記憶システムの第三の制御装置は、所定のジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第三のジャーナル記憶領域に書くようになっている。

[0024]

第二の記憶システムは、データの複製を格納する第二のデータ記憶領域を更に有することができる。第二の制御装置は、所定のリストアのタイミングで第二のジャーナル記憶領域に格納されたジャーナルからデータの複製を生成し、生成されたデータの複製を第二のデータ記憶装置に書くようになっていてよい。

[0025]

第三の記憶システムも、データの複製を格納する第三のデータ記憶領域を更に有することができる。第三の制御装置は、所定のリストアのタイミングで第三のジャーナル記憶領域に格納されたジャーナルからデータの複製を生成し、生成されたデータの複製を第三のデータ記憶装置に書くようになっていてよい。

[0026]

第一の記憶システムの第一の制御装置は、第一のジャーナル記憶領域内のジャーナルが 第二及び第三の記憶システムによって読まれたか否かを検出し、第二及び第三の記憶シス テムの双方によって読まれるまでは第一のジャーナル記憶領域内の前記ジャーナルを保持 し、第二及び第三の記憶システムの双方によって読まれた後に第一のジャーナル記憶領域 内のジャーナルを消去可能とするようになっていてよい。

【発明を実施するための最良の形態】

[0027]

本発明によるデータ処理システムの実施形態を図面により詳細に説明する。

[0028]

図1は、本発明に従うデータ処理システムの一実施形態の物理的な構成を示すブロック 図である。図2は、本実施形態の論理的な構成を示すブロック図である。

[0029]

図1及び図2に示すように、このデータ処理システムでは、ホストコンピュータ180と記憶システム100Aが接続パス190により接続され、また、記憶システム100Aと別の記憶システム100Bとが接続パス200により接続される。記憶システム100Bは、記憶システム100Aに保存されたデータの複製を保持するために使用される。以下の説明において、複製対象のデータを保持する記憶システム100Aと複製データを保持する記憶システム100Bとの区別を容易とするために、前者100Aを「正記憶システム」、後者100Bを「副記憶システム」と呼ぶこととする。

[0030]

正記憶システム100Aと副記憶システム100Bの物理的構成について、図1を参照して説明する。

$[0\ 0\ 3\ 1]$

正記憶システム100Aと副記憶システム100Bは基本的には同様の物理的構成を有するので、図1では、代表的に正記憶システム100Aの構成のみが示されている。図1に示すように、正記憶システム100Aは、1つ以上のチャネルアダプタ110、1つ以上のディスクアダプタ120、1つ以上のキャッシュメモリ130、1つ以上の共有メモリ140、1つ以上の物理的記憶装置(例えば、ハードディスクドライブ)150、1つ以上のコモンパス160、1つ以上の接続線170を備える。チャネルアダプタ110、ディスクアダプタ120、キャッシュメモリ130、共有メモリ140はコモンパス160により相互に接続されている。コモンパス160は、コモンパス160の障害時のために2重化されてもよい。ディスクアダプタ120と物理的記憶装置150とは接続線170によって接続されている。また、図示していないが、記憶システム100の設定、監視

5/

、保守等を行うための保守端末が全てのチャネルアダプタ110とディスクアダプタ12 0とに専用線を用いて接続されている。

[0032]

チャネルアダプタ110は、接続線190又は接続線200によりホストコンピュータ180又は他の記憶システム(例えば副記憶システム100B)と接続される。チャネルアダプタ110は、ホストコンピュータ180とキャッシュメモリ130間のデータ転送、又は他の記憶システムとキャッシュメモリ130間のデータ転送を制御する。ディスクアダプタ120は、キャッシュメモリ130と物理的記憶装置150との間のデータ転送を制御する。キャッシュメモリ130は、ホストコンピュータ180又は他の記憶システムから受信したデータ、あるいは物理的記憶装置150から読み出したデータを一時的に保持するメモリである。共有メモリ140は、記憶システム100内の全てのチャネルアダプタ110とディスクアダプタ12が使用する制御や管理のための様々な情報(例えば、後述するボリューム情報400、ペア情報500、グループ情報600およびポインタ情報700など)が記憶され保持される。副記憶システム100Bの物理的構成も、基本的にこれと同様である。

[0033]

次に、記憶システム100A及び100Bの論理的な構成について、図2を参照して説明する。

$[0\ 0\ 3\ 4]$

図2に示すように、記憶システム100A及び100Bの各々において、物理的記憶装 置150、150、…により提供される全体の記憶領域は、論理的な多数の記憶領域23 0、230、…に分割されて管理される。以下、個々の論理的な記憶領域230を「論理 ボリューム」と呼ぶこととする。各論理ボリューム230の容量および記憶システム10 0 A 又は100 B 内での物理的な格納位置(物理アドレス)は、記憶システム100 A 又 は100Bに接続された保守用の端末コンピュータ(図示せず)もしくはホストコンピュ ータ180から指定することができる。各論理ボリューム230の物理アドレスは、後述 するボリューム情報400に保存される。物理アドレスは、例えば、記憶システム100 A又は100B内の物理的記憶装置150を識別する番号(記憶装置番号)と、その記憶 装置150内での記憶領域を一意に示す数値(例えば、物理的記憶装置150内での記憶 領域の先頭からの位置)から構成される。以下の説明では、物理アドレスは、物理的記憶 装置150の記憶装置番号とその物理的記憶装置150内での記憶領域の先頭からの位置 の組で表されるものとする。普及しているRAIDの原理に従った記憶システムでは、1 つの論理ボリューム230が複数の物理的記憶装置150内の複数の物理的な記憶領域に 対応しているが、以下では、説明を容易にするために、1つの論理ボリューム230が、 1つの物理的な記憶装置150内の1つの記憶領域に対応するものとして説明する。しか し、RAIDの原理に従う記憶システムでも本発明の原理が以下の説明と同様に適用でき ることを、当業者は容易に理解する筈である。

[0035]

記憶システム100A、100B内に保存されているデータは、そのデータが存在する 論理ボリューム230を識別する番号(論理ボリューム番号)と、その論理ボリューム2 030内のそのデータの記憶領域を一意に示す数値(例えば、論理ボリューム230の記 憶領域の先頭からの位置)により一意に指定することができる。以下の説明では、論理ア ドレスは、論理ボリューム230の論理ボリューム番号と論理ボリューム230内での記 憶領域の先頭からの位置(論理ボリューム内位置)の組で表されるものとする。

[0036]

以下の説明において、複製対象の論理ボリュームとその複製である論理ボリュームとの区別を容易とするために、前者を「正論理ボリューム」、後者を「副論理ボリューム」とよぶこととする。一対の正論理ボリュームと副論理ボリュームを「ペア」とよぶこととする。正論理ボリュームと副論理ボリュームの関係および状態等の情報は後述するペア情報

500に保存される。

[0037]

複数の正論理ボリュームと、それらの正論理ボリュームとペアをそれぞれ構成する複数 の副論理ボリュームとの間のデータの更新順序を一致させるために、「グループ」という 管理単位が設けられる。例えば、ホストコンピュータ180が、第1の正論理ボリューム 内の第1のデータを更新し、その後、その第1のデータを読み出し、その第1のデータを 用いて、第2の正論理ボリューム内の第2のデータを更新する処理を行う場合を想定する 。この場合、第1の正論理ボリュームから第1の副論理ボリュームへのデータ複製処理と 、第2の正論理ボリュームから第2の副論理ボリュームへのデータ複製処理とが独立に行 われたならば、第1の副論理ボリュームへの第1のデータの複製処理より前に、第2の副 論理ボリュームへの第2のデータの複製処理が行われる場合がある。この場合に、もし、 第2の副論理ボリュームへの第2のデータの複製処理の終了後、第1の副論理ボリューム への第1のデータの複製処理が終了する前に、故障等が発生して、第1の副論理ボリュー ムへの複製処理が停止したならば、第1の副論理ボリュームと第2の副論理ボリュームの 間でのデータの整合性が損なわれることになる。このような場合であって、第1の副論理 ボリュームと第2の副論理ボリューム間のデータの整合性を保つために、「グループ」を 用いた更新順序の制御が行われる。すなわち、正論理ボリュームと副論理ボリューム間で データの更新順序を一致させる必要のある複数の論理ボリュームが、同じグループに登録 される。そのグループに属する正論理ボリュームで行われるデータの更新毎に、後述する グループ情報600の更新番号が割り当てられる。そして、その更新番号順に、更新され たデータの副論理ボリュームへの複製処理が行われる。例えば、図2の例では、正記憶シ ステム100A内の2つの正論理ボリューム「DATA1|と「DATA2|が一つのグ ループ「グループ1」を構成する。そして、これら2つの正論理ボリューム「DATA1 」と「DATA2」の複製である2つの副論理ボリューム「COPY1」と「COPY2 」が、副記憶システム100B内で同じグループ「グループ1」を構成する。

[0038]

データ複製対象である正論理ボリュームのデータを更新する場合、副論理ボリュームのデータの更新に利用するために、そのデータ更新についてのジャーナルが作成され、そのジャーナルが正記憶システム100A内の所定の論理ボリュームに保存される。本実施形態では、正記憶システム110A内の各グループに対して、そのグループのジャーナルのみを保存するための論理ボリューム(以下、ジャーナル論理ボリュームとよぶ)が割り当てられる。図2の例では、グループ「グループ1」に、ジャーナル論理ボリューム「JNL1」が割り当てられている。副記憶システム100B内の各グループにも、ジャーナル論理ボリュームが割り当てられる。図1では、副記憶システム内のグループ「グループ1」に、ジャーナル論理ボリューム「JNL2」が割り当てられている。副記憶システム100Aから副記憶システム100B内のジャーナルを保存するために使用ことにより、ジャーナルを副記憶システム100B内のジャーナル論理ボリュームに保存することができる。ジャーナルに基づいて副論理ボリュームのデータを更新することができる。

[0039]

さらに、正記憶システム100Aと副記憶システム100Bとの間の接続線200が複数ある場合、正記憶システム100Aから副記憶システム100Bへのジャーナルの転送を多重に行い、複数の接続線200の転送能力を有効に利用することができる。所定の更新順序を守ることにより、副記憶システム100Bに多くのジャーナルが溜まる可能性があるが、副論理ボリュームのデータ更新に直ぐに使用できないジャーナルは、ジャーナル論理ボリュームに退避することにより、キャッシュメモリを開放することができる。

[0040]

上述のジャーナルは、ライトデータと更新情報とから構成される。更新情報は、ライト

7/

データを管理するための情報で、ライト命令を受信した時刻、グループ番号、後述するグループ情報600の更新番号、ライト命令の論理アドレス、ライトデータのデータサイズ、ライトデータを格納したジャーナル論理ボリュームの論理アドレス等を含む。更新情報は、ライト命令を受信した時刻と更新番号のどちらか一方のみを保持してもよい。ホストコンピュータ180からのライト命令の中にライト命令の作成時刻が含まれている場合は、ライト命令を受信した時刻の代わりに、当該ライト命令内の命令作成時刻を使用してもよい。

$[0\ 0\ 4\ 1]$

図3はジャーナルの更新情報とライトデータの関係の一例を示す。

[0042]

図3に示すように、ジャーナル論理ボリューム350は、例えば、更新情報を格納する 記憶領域(更新情報領域)351と、ライトデータを格納する記憶領域(ライトデータ領 域)353に分割されて使用される。正論理ボリューム34に書き込まれたライトデータ 3 2 0 の更新情報 3 1 0 は、ジャーナル論理ボリューム 3 5 0 の更新情報領域 3 5 1 の先 頭から、更新番号の順に格納される。或る更新情報310の格納場所が更新情報領域35 1の終端に達すると、後続の更新情報310は更新情報領域351の先頭から格納される 。正論理ボリューム34に書き込まれたライトデータ320に対応するライトデータ33 0が、ジャーナル論理ボリューム350のライトデータ領域353先頭から、更新番号の 順に格納される。或るライトデータ330の格納場所がライトデータ領域353の終端に 達すると、後続のライトデータ330はライトデータ領域351の先頭から格納される。 更新情報領域351およびライトデータ領域353のサイズ比は、固定値でもよいし、保 守端末あるいはホストコンピュータ180により任意値に設定可能であってもよい。ジャ ーナル論理ボリューム350内での更新情報領域351およびライトデータ領域353の 範囲を示したアドレスや、最新及び最古の更新情報310およびライトデータ330の記 憶場所を示したアドレスなどの情報は、後述するポインタ情報700内に保持される。以 下の説明では、ジャーナル論理ボリューム350が更新情報領域351とライトデータ領 域353に分割されて使用されるが、変形例として、ジャーナル論理ボリューム350の 先頭から、各ジャーナルの更新情報310とライトデータ330を連続的に格納される方 式を採用してもよい。

[0043]

図4は、ジャーナルの更新情報310の具体例を示す。

[0044]

$[0\ 0\ 4\ 5]$

再び図2を参照する。正記憶システム100Aは、ホストコンピュータ180からのデータのリード/ライト命令を受付ける命令受付処理210、及びデータを然るべき論理ボリューム230にリード/ライトするリードライト処理220を行う機能を有する。副記憶システム100Bは、正記憶システム100Aからジャーナルを読むジャーナルリード(JNLRD)処理240、データを然るべき論理ボリューム230にリード/ライトするリードライト処理220、及び正記憶システム100Aからのジャーナルに基づいて然るべき副論理ボリューム230内のデータを更新するリストア処理250を行う機能を有

する。これらの処理機能は、図2に示したチャネルアダプタ110、ディスクアダプタ1 20、キャッシュメモリ130及び共有メモリ140によって実現される。

[0046]

図2を参照して、正記憶システム100Aの正論理ボリュームへのデータ更新を副記憶 システム100Bの副論理ボリュームに反映する動作について概説する。

[0047]

(1) 正記憶システム 100 A は、ホストコンピュータ 180 から正論理ボリューム (例えば「DATA1」) 230 内のデータに対するライト命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、正論理ボリューム(「DATA1」) 230 内の対象データの更新と、ジャーナル論理ボリューム(「JNL1」) 230 へのジャーナルの保存とを行う(矢印 270)。

[0048]

(2) 副記憶システム100Bは、ジャーナルリード処理240によって、正記憶システム100Aにジャーナルをリードする命令を送る。正記憶システム100Aは、副記憶システム100Bからジャーナルをリードする命令を受信すると、命令受信処理210およびリードライト処理220によって、ジャーナル論理ボリューム(「JNL1」)230からジャーナルを読み出し、副記憶システム100Bに送信する(矢印280)。

[0049]

(3) 副記憶システム 100B は、正記憶システムからリードしたジャーナルを、リードライト処理 220 によって、ジャーナル論理ボリューム(「JNL2」) 230 に保存する(矢印 280)。

[0050]

(4) 副記憶システム100Bは、リストア処理250およびリードライト処理22 0によって、ポインタ情報700を用いて、更新番号の昇順に、ジャーナル論理ボリューム(「JNL2」)230からジャーナルを読み出し、副論理ボリューム(例えば「COPY1」)230のデータを更新する(矢印290)。

[0051]

図5は、ボリューム情報400の具体例を示す。

[0052]

ボリューム情報400は、チャネルアダプタ110およびディスクアダプタ120から 参照可能なメモリ、例えば図1に示した共有メモリ140に保存される。ボリューム情報 400は、論理ボリュームを管理するための情報であり、図5に示すように、論理ボリュ ーム番号毎に、ボリューム状態、フォーマット形式、容量、ペア番号、物理アドレスを保 持する。ボリューム状態は、"正常"、"正"、"副"、"異常"、"未使用"のいずれ かを保持する。ボリューム状態が"正常"もしくは"正"である論理ボリューム230は 、ホストコンピュータ180から正常にアクセス可能な論理ボリューム230である。ボ リューム状態が"副"である論理ボリューム230は、ホストコンピュータ180からの アクセスを許可してもよい。ボリューム状態が"正"である論理ボリューム230は、デ ータの複製が行われている正論理ボリューム230である。ボリューム状態が"副"であ る論理ボリューム230は、複製に使用されている副論理ボリューム230である。ボリ ューム状態が"異常"の論理ボリューム230は、障害により正常にアクセスできない論 理ボリューム230である。ここで、障害とは、例えば、論理ボリューム230を保持す る記憶装置 150の故障である。ボリューム状態が"未使用"の論理ボリューム 230は 、使用していない論理ボリューム230である。ペア番号は、ボリューム状態が"正"も しくは"副"の場合に有効であり、後述するペア情報500を特定するためのペア番号を 保持する。図5に示す例では、論理ボリューム番号「1」の論理ボリュームは、複製対象 のデータを保持した正論理ボリュームであり、アクセス可能であり、フォーマット形式が | OPEN3」であり、容量が「3GB」であり、そこに保持されたデータは、記憶装置 番号「1」の物理的記憶装置150の記憶領域の先頭から格納されていることが、示され ている。

[0053]

図6は、ペア情報500の具体例を示す。

[0054]

ペア情報500は、チャネルアダプタ110およびディスクアダプタ120から参照可 能なメモリ、例えば図1に示した共有メモリ140に保存される。ペア情報500は、ペ アを管理するための情報であり、図6に示すように、ペア番号毎に、ペア状態、正記憶シ ステム番号、正論理ボリューム番号、副記憶システム番号、副論理ボリューム番号、グル ープ番号、コピー済みアドレスを保持する。ペア状態は、"正常"、"異常"、"未使用 ","コピー未"、"コピー中"のいずれかを保持する。ペア状態が"正常"の場合は、 正論理ボリューム230のデータ複製が正常に行われていることを示す。ペア状態が"異 常"の場合は、障害により正論理ボリューム230の複製が行えないことを示す。ここで 、障害とは、例えば、接続パス200の断線などである。ペア状態が"未使用"の場合は 、当該ペア番号の情報は有効でないことを示す。ペア状態が"コピー中"の場合は、後述 する初期コピー処理中であることを示す。ペア状態が"コピー未"の場合は、後述する初 期コピー処理が未だ行われていないことを示す。正記憶システム番号は、正論理ボリュー ム230を保持する正記憶システム100Aを特定する番号を保持する。副記憶システム 番号は、副論理ボリューム230を保持する副記憶システム100Bを特定する番号を保 持する。グループ番号は、正記憶システム100Aの場合は、正論理ボリュームが属する グループ番号を保持し、副記憶システム100Bの場合は、副論理ボリュームが属するグ ループ番号を保持する。コピー済みアドレスは、後述する初期コピー処理にて説明する。 図6に例示したペア情報1は、データ複製対象が正記憶システム番号「1 | の正論理ボリ ューム番号「1 | により特定される正論理ボリュームであり、データ複製先が副記憶シス テム番号「2」の副論理ボリューム番号「1」で特定される副論理ボリュームであり、正 常にデータ複製処理が行われていることを示す。

[0055]

図7は、グループ情報600の具体例を示す。

[0056]

グループ情報600は、チャネルアダプタ110およびディスクアダプタ120から参 照可能なメモリ、例えば共有メモリ140に保存される。図7に示すように、グループ情 報600は、グループ番号毎に、グループ状態、ペア集合、ジャーナル論理ボリューム番 号、更新番号を保持する。グループ状態は、"正常"、"異常"、"未使用"のいずれか を保持する。グループ状態が"正常"の場合は、ペア集合の少なくともひとつのペア状態 が"正常"であることを示す。グループ状態が"異常"の場合は、ペア集合の全てのペア 状態が"異常"であることを示す。グループ状態が"未使用"の場合は、当該グループ番 号の情報は有効でないことを示す。ペア集合は、正記憶システムの場合は、そのグループ に属する全ての正論理ボリュームのペア番号を保持し、副記憶システムの場合は、そのグ ループに属する全ての副論理ボリュームのペア番号を保持する。ジャーナル論理ボリュー ム番号は、当該グループ番号のグループに属するジャーナル論理ボリューム番号を示す。 更新番号は、初期値は1であり、グループ内の正論理ボリュームに対しデータの書き込み が行われると、+1だけ増加する。更新番号は、ジャーナルの更新情報に記憶され、副記 憶システム100Bにて、データの更新順を守るために使用する。図7に零時さえタグル ープ情報600は、グループ番号「1」のグループが、ペア番号「1」及び「2」に属す る論理ボリュームと、論理ボリューム番号「4」のジャーナル論理ボリューム4から構成 されており、正常にデータの複製処理が行われていることを示す。

[0057]

図8は、ポインタ情報700の具体例を示す。図9は、ポインタ情報に含まれる項目と ジャーナル論理ボリューム350との関係を説明した図である。

[0058]

ポインタ情報700は、チャネルアダプタ110およびディスクアダプタ120から参照可能なメモリ、例えば共有メモリ140に保存される。ポインタ情報700は、グルー

プ毎に、当該グループのジャーナル論理ボリュームを管理するための情報であり、図8に示すように、そのジャーナル論理ボリューム内の更新情報領域先頭アドレス、ライトデータ領域先頭アドレス、更新情報最新アドレス、更新情報最古アドレス、ライトデータ最新アドレス、ライトデータ最古アドレス、リード開始アドレス及びリトライ開始アドレスを保持する。

[0059]

正記憶システム100Aは、正記憶システム100A内のジャーナル論理ボリュームを 管理するためのポインタ情報700を、正記憶システム100Aに接続された副記憶シス テムの台数分のセット数だけ有する。すなわち、図2に示した例では、正記憶システム1 00Aに接続された副記憶システム100Bは1台であるため、図8に示すように、正記 憶システム100Aは、その一つの副記憶システム100Bに対応した1セットのポイン 夕情報700を有し、そのポインタ情報700には、その一つの副記憶システム100B のシステム番号が記載されている。しかし、後に説明する図25の例では、正記憶システ ム100Aに複数(例えば2台)の副記憶システム100B、100Cが並列に接続され る。図25の例の場合、正記憶システム100Aは、それら複数の副記憶システム100 B、100Cにそれぞれに対応した複数セットのポインタ情報700B、700C(図2 6参照)を有し、それらの複数セットのポインタ情報700B、700Cには対応する副 記憶システム100B、100Cのシステム番号が記録されている。このように正記憶シ ステム100Aが、それに接続された複数の副記憶システム100B、100Cにそれぞ れ対応する複数のポインタ情報700を有する場合、それぞれの副記憶システム100B 、100Cによる正記憶システム100Aからのジャーナルリードが行われたか否か(そ のタイミングは副記憶システム100B、100Cによって異なる)が、副記憶システム 100B、100Cにそれぞれ割り当てられた複数のポインタ情報700によって管理す ることができる。これを利用して、正記憶システム100Aは、正記憶システム100A 内の各ジャーナルを、それが複数の副記憶システム100B、100Cの全てによってリ ードされ終わらない限り、消去されないよう保持するよう制御を行うことができる。

[0060]

一方、副記憶システム100Bは、基本的に、その副記憶システム110B内のジャーナル論理ボリュームを管理するための1セットのポインタ情報700を有する。ただし、特に図示されてはいないが、正記憶システムに複数の副記憶システムがカスケード接続された構成も採用でき、この場合、そのカスケードの中間に位置する副記憶システムは、その上側と下側の2つの副記憶システムにそれぞれ対応した複数セットのポインタ情報をもつことができる。

$[0\ 0\ 6\ 1]$

図9に示すように、更新情報領域先頭アドレスは、ジャーナル論理ボリューム350の 更新情報領域353の先頭の論理アドレスを保持する。ライトデータ領域先頭アドレスは 、ジャーナル論理ボリューム350のライトデータ領域353も先頭の論理アドレスを保 持する。更新情報最新アドレスは、次にジャーナルを格納する場合に、その更新情報の保 存に使用される先頭の論理アドレスを保持する。更新情報最古アドレスは、最古の(更新 番号が小さい)ジャーナルの更新情報を保存する先頭の論理アドレスを保持する。ライト データ最新アドレスは、次にジャーナルを格納する場合に、ライトデータの保存に使用す る先頭の論理アドレスを保持する。ライトデータ最古アドレスは、ジャーナル論理ボリュ ーム350内での最古の(更新番号が最も小さい)ジャーナルのライトデータを保存する 先頭の論理アドレスを保持する。リード開始アドレスとリトライ開始アドレスは、正記憶 システム100Aのみで使用され、後述するジャーナルリード受信処理にて使用される。 図8に例示されたポインタ情報700によれば、更新情報領域35が論理ボリュー番号「 4」の記憶領域の先頭(アドレス「0」)からアドレス「699」の位置までであり、ラ イトデータ領域353が論理ボリューム番号「4」の記憶領域のアドレス「700」の位 置からアドレス「2699」の位置までである。そして、更新情報は、論理ボリューム番 号「4」の記憶領域のアドレス「200」の位置からアドレス「499」の位置まで保存 されており、次のジャーナルの更新情報は、論理ボリューム番号「4」の記憶領域のアドレス「500」の位置から格納されることになる。また、ジャーナルのライトデータは、論理ボリューム番号「4」の記憶領域のアドレス「1300」の位置からアドレス「2199」の位置までに保存されており、次のジャーナルのライトデータは、論理ボリューム番号「4」の記憶領域のアドレス「2200」の位置から格納されることになる。

[0062]

ここでの説明では、1つのグループに1つのジャーナル論理ボリュームが割り当てられる。しかし、1つのグループに複数のジャーナル論理ボリュームを割り当てられてもよい。例えば、1つのグループに2つのジャーナル論理ボリュームが割り当てられ、ジャーナル論理ボリューム毎にポインタ情報700が設けられ、そして、その2つのジャーナル論理ボリュームに交互にジャーナルが格納されるようにしてもよい。これにより、ジャーナルの物理的記憶装置150への書き込みが分散され、性能の向上が見込める。さらに、ジャーナルのリード性能も向上する。別の例としては、1つのグループに2つのジャーナル論理ボリュームが割り当てられ、通常は、1つのジャーナル論理ボリュームのみが使用される。もう一方のジャーナル論理ボリュームは、現在使用されてきたジャーナル論理ボリュームの性能が低下した場合に使用される。性能が低下する場合とは、例えば、ジャーナル論理ボリュームが、複数の物理的記憶装置150から構成され、150の一台が故障した場合である。

[0063]

上述したように、ボリューム情報400、ペア情報500、グループ情報600、及びポインタ情報7.00等は、共有メモリ140に格納されることができる。しかし、これらの情報400、500、600及び700が、キャッシュメモリ130、チャネルアダプタ110、ディスクアダプタ120、又は物理的記憶装置150のいずれかに集中して、又はこれらに分散して格納されてもよい。

[0064]

図10は、本実施形態においてデータの複製を開始する手順を説明するフローチャートである。図10を参照して、正記憶システム100Aから副記憶システム100Bに対して、データ複製を開始する手順を説明する。

[0065]

(1) グループ作成 (ステップ900)

ユーザは、保守端末あるいはホストコンピュータ180を使用して、正記憶システム100A内のグループ情報600を参照し、グループ状態が"未使用"のグループ番号、例えば「A」、を取得する。ユーザは、保守端末あるいはホストコンピュータ180を使用して、そのグループ番号「A」を指定したグループ作成指示を、正記憶システム100Aに入力する。グループ作成指示を受けて、正記憶システム100Aは、指定されたグループ番号「A」のグループ状態を"正常"に変更する。

[0066]

同様に、ユーザは、副記憶システム100Bのグループ情報600を参照し、グループ 状態が"未使用"のグループ番号、例えば「B」、を取得する。ユーザは、保守端末ある いはホストコンピュータ180を使用して、副記憶システム100Bとグループ番号「B 」を指定したグループ作成指示を、正記憶システム100Aに入力する。正記憶システム 100Aは、受信したグループ作成指示を副記憶システム100Bに転送する。副記憶システム100Bは、指定されたグループ番号「B」のグループ状態を"正常"に変更する

[0067]

或いは、ユーザは、副記憶システム100Bの保守端末あるいは副記憶システム100Bに接続されたホストコンピュータ180を使用して、グループ番号「B」を指定したグループ作成指示を、副記憶システム100Bに直接入力してもよい。この場合も、副記憶システム100Bは、指定されたグループ番号「B」のグループ状態を"正常"に変更す

る。

[0068]

(2) ペア登録 (ステップ910)

ユーザは、保守端末あるいはホストコンピュータ180を使用して、データ複製元である正論理ボリュームを特定する情報とデータ複製先である副論理ボリュームを特定する情報を指定したペア登録指示を、正記憶システム100Aに入力する。正論理ボリュームを特定する情報には、作成されたグループ番号「A」と、そのグループ「A」内の正論理ボリューム番号が含まれる。副論理ボリュームを特定する情報には、副記憶システム100Bの記憶システム番号と、作成されたグループ番号「B」と、そのグループ「B」内の副論理ボリューム番号が含まれる。

[0069]

前記ペア登録指示を受けて、正記憶システム100Aは、ペア情報500からペア情報が"未使用"のペア番号を取得する。そして、正記憶システム100A、ペア情報500内の取得したペア番号の行において、ペア状態を"コピー未"に設定し、正記憶システム番号に、正記憶システム番号に、正記憶システム番号を設定し、正論理ボリューム番号に、指示された正論理ボリューム番号を設定し、副記憶システム番号に、指示された副記憶システム番号を設定し、副論理ボリューム番号に、指示された副論理ボリューム番号を設定し、そして、グループ番号に、指示されたグループ番号「A」を設定する。さらに、正記憶システム100Aは、指示されたグループ番号「A」のグループ情報600のペア集合に、前記取得したペア番号を追加し、正論理ボリューム番号のボリューム状態を"正"に変更する。

[0070]

正記憶システム100Aは、正記憶システム100Aの記憶システム番号、ユーザから指定されたグループ番号「B」、正論理ボリューム番号、および副論理ボリューム番号を、副記憶システム100Bは、ペア情報500から未使用のペア番号を取得する。そして、副記憶システム100Bは、ペア情報500内の取得したペア番号の行において、ペア状態を"コピー未"に設定し、正記憶システム番号に、記憶システム100Aの記憶システム番号を設定し、正論理ボリューム番号に、指示された正論理ボリューム番号を設定し、副記憶システム看号に、副記憶システム100Bの副記憶システム番号を設定し、副論理ボリューム番号に、指示されたグループ番号「B」を設定する。さらに、副記憶システム100Bは、指示されたグループ番号「B」を設定する。さらに、副記憶システム100Bは、指示されたグループ番号「B」のグループ情報600のペア集合に、前記取得したペア番号を追加し、副論理ボリューム番号のボリューム状態を"副"に変更する。

[0071]

以上の動作が、全てのデータ複製対象のペアに対して行われる。

[0072]

前記の説明では、論理ボリュームのグループへの登録と、論理ボリュームのペアの設定 を同時に行う処理を説明したが、それぞれ個別に行ってもよい。

[0073]

(3) ジャーナル論理ボリューム登録(ステップ920)

ユーザは、保守端末あるいはホストコンピュータ180を使用して、ジャーナルの保存に使用する論理ボリューム (ジャーナル論理ボリューム) をグループに登録する指示 (ジャーナル論理ボリューム登録指示) を、正記憶システム100Aに入力する。ジャーナル論理ボリューム登録指示には、グループ番号と論理ボリューム番号が含まれる。

[0074]

正記憶システム100Aは、指示されたグループ番号のグループ情報600内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報400内のボリューム状態に、"正常"を設定する。

[0075]

同様に、ユーザは、保守端末あるいはホストコンピュータ180を使用して、副記憶システム100Bのボリューム情報400を参照し、副記憶システム100Bの記憶システム番号、グループ番号「B」、ジャーナル論理ボリュームの論理ボリューム番号を指定して、ジャーナル論理ボリューム登録の指示を正記憶システム100Aに入力する。正記憶システム100Aは、そのジャーナル論理ボリューム登録指示を副記憶システム100Bに転送する。副記憶システム100Bは、指示されたグループ番号「B」のグループ情報600内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報400内のボリューム状態に、"正常"を設定する。

[0076]

ユーザは、副記憶システム100Bの保守端末あるいは副記憶システム100Bに接続したホストコンピュータ180を使用して、グループ番号、ジャーナル論理ボリュームの論理ボリューム番号を指定して、ジャーナル論理ボリューム登録指示を副記憶システム100Bに直接入力してもよい。この場合も、副記憶システム100Bは、指示されたグループ番号「B」のグループ情報600内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報400内のボリューム状態に、"正常"を設定する。

[0077]

以上の動作が、ジャーナル論理ボリュームとして使用される全ての論理ボリュームに関して行われる。ステップ910とステップ920の順は逆であってもよい。

[0078]

(4) データ複製処理の開始 (ステップ930)

ユーザは、保守端末あるいはホストコンピュータ180を使用して、データ複製処理を開始するグループ番号を指定して、データ複製処理の開始を正記憶システム100Aに指示する。正記憶システム100Aは、指示されたグループに属する全てのペア情報400のコピー済みアドレスを"0"に設定する。

[0079]

正記憶システム100Aは、副記憶システム100Bに、ジャーナルリード処理240およびリストア処理250(図2参照)の開始を指示する。正記憶システム100Aは、後述する初期コピー処理を開始する。

[0080]

(5) 初期コピー処理の終了 (ステップ940)

初期コピー処理が終了すると、正記憶システム100Aは、初期コピー処理の終了を副記憶システム100Bに通知する。副記憶システム100Bは、指示されたグループに属する全ての副論理ボリュームのペア状態を"正常"に変更する。

[0081]

図11は、上述した図10のステップ930で行われる初期コピー処理の手順を示すフローチャートである。

[0082]

初期コピー処理では、データ複製対象の正論理ボリュームの全記憶領域に関して、ペア情報500のコピー済みアドレスを用いながら、その全記憶領域の先頭から順に、単位データサイズ毎にジャーナルが作成される。コピー済みアドレスは、初期値は0であり、そこに、ジャーナルが作成される都度、ジャーナル作成が済んだデータのデータサイズが加算される。論理ボリュームの記憶領域の先頭から、コピー済みアドレスの直前のアドレスまでの範囲については、初期コピー処理にてジャーナルが作成済みである。初期コピー処理を行うことにより、正論理ボリューム内の更新されていないデータを副論理ボリュームに転送することが可能となる。以下の説明では、正記憶システム100A内のチャネルアダプタ110(図1参照)が初期コピー処理を行うように説明されるが、代わって、ディスクアダプタ120がそれを行ってもよい。以下、図11に基づいて初期コピー処理の手順を説明する。

[0083]

(1) 正記憶システム 100 A内のチャネルアダプタ 110は、処理対象のグループに属するペアの中でペア状態が"コピー未"である正論理ボリューム、例えば「A」を得、そのペアの状態を"コピー中"に変更し、以下の処理を繰り返す(ステップ 1010、1020)。もし、"コピー未"のペア状態にある正論理ボリュームが存在しない場合は、処理を終了する(ステップ 1030)。

[0084]

(2) ステップ1010にて取得した正論理ボリューム「A」について、チャネルアダプタ110は、単位サイズ (例えば、1MB) のデータ毎に、ジャーナルを作成する (ステップ1040)。ジャーナル作成処理の詳細については後述する

[0085]

(3) チャネルアダプタ110は、ペア情報500のコピー済みアドレスに、ジャーナル作成が済んだデータのサイズを加算する(ステップ1050)。

[0086]

(4) コピー済みアドレスが、正論理ボリューム「A」の容量に達するまで、上記ステップ10.40及び1050の処理を繰り返す(ステップ1060)。コピー済みアドレスが、正論理ボリューム「A」の容量と等しくなった場合、正論理ボリューム「A」の全記憶領域に対してジャーナルが作成されたことになるため、チャネルアダプタ110は、ペア状態を"正常"に更新し、そして、他の正論理ボリュームについて初期コピー処理を開始する(ステップ1070)。

[0087]

図11のフローチャートでは、個々の正論理ボリュームが逐次に処理されるが、複数の 論理ボリュームが同時に処理されてもよい。

[0088]

図12は、命令受信処理210(図2参照)におけるデータの流れを説明する図である。図13は、命令受信処理210の手順を示すフローチャートである。図14は、ジャーナル作成処理のフローチャートである。以下、図12~図14を参照して、正記憶システム100Aが、ホストコンピュータ180からデータ複製対象の論理ボリューム230にライト命令を受信した場合の動作について説明する。

[0089]

まず、図12及び図13を参照して、命令受信処理210の全体的な動作を説明する。

[0090]

(1) 正記憶システム100A内のチャネルアダプタ110は(図13のステップ1200)、ホストコンピュータからアクセス命令を受信する。アクセス命令は、データリード、データライト、又は後述するジャーナルリード等の命令、命令対象の論理アドレス、及びデータ量等を含んでいる。以下の説明では、アクセス命令内の論理アドレスを論理アドレス「A」、論理ボリューム内位置を論理ボリューム内位置「A」、データ量をデータ量「A」とする。

[0091]

(2) チャネルアダプタ110は、アクセス命令を調べる(ステップ1210、1215)。ステップ1215の調べで、アクセス命令がジャーナルリード命令の場合は、後述するジャーナルリード受信処理を行う(ステップ1220)。アクセス命令がジャーナルリード命令およびデータライト命令以外、例えば、データリード命令の場合は、従来技術と同じようにデータリード処理を行う(ステップ1230)。

[0092]

(3) ステップ1210の調べで、アクセス命令がデータライト命令の場合は、チャネルアダプタ110は、論理ボリュームAのボリューム情報400を参照し、ボリューム状態を調べる(ステップ1240)。ステップ1240の調べで、論理ボリュームAのボリューム状態が、"正常"もしくは"正"以外の場合は、論理ボリュームAへのアクセスは不可能なため、チャネルアダプタ110は、ホストコンピュータ180に異常終了を報

告する(ステップ1245)。

[0093]

(4) ステップ1240の調べで、論理ボリュームAのボリューム状態が、"正常"、"正"のいずれかの場合は、チャネルアダプタ110は、キャッシュメモリ130を確保し、ホストコンピュータ180にデータ受信の準備ができたことを通知する。ホストコンピュータ180は、その通知を受け、ライトデータを正記憶システム100Aに送信する。チャネルアダプタ110は、ライトデータを受信し、これを当該キャッシュメモリ130に保存する(ステップ1250、図12の1100)。

[0094]

(5) チャネルアダプタ 1 1 0 は、論理ボリューム「A」のボリューム状態を参照し、論理ボリューム「A」がデータ複製対象かどうかを調べる(ステップ 1 2 6 0 0 。ステップ 1 2 6 0 の調べで、ボリューム状態が、 "正"である場合は、論理ボリューム「A」がデータ複製対象であるため、後述するジャーナル作成処理を行う(ステップ 1 2 6 5 0

[0095]

(6) ステップ1260の調べで、ボリューム状態が、"正常"である場合、もしくはステップ1265のジャーナル作成処理の終了後、チャネルアダプタ110は、ディスクアダプタ120にライトデータを記憶装置150に書き込むことを命令し(図12の1140)、ホストコンピュータ180に終了報告する(ステップ1270、1280)。その後、当該ディスクアダプタ120は、リードライト処理により、物理的記憶装置150にライトデータを保存する(図12の1110)。

[0096]

次に、上述した図11のステップ1040又は図13のステップ1265のジャーナル 作成処理について、図12及び図14を参照して説明する。

[0097]

(1) チャネルアダプタ110は、最初のセットのポインタ情報700を取得する(図14のステップ1305)。チャネルアダプタ110は、ジャーナル論理ボリュームのボリューム状態を調べる(ステップ1310)。ステップ1310の調べで、ジャーナル論理ボリュームのボリューム状態が、 "異常"の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、チャネルアダプタ110は、グループ状態を "異常"に変更し、処理を終了する(ステップ1315)。この場合、チャネルアダプタ110は、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行う。

[0098]

(2) ステップ1310の調べで、ジャーナル論理ボリュームが正常である場合、チャネルアダプ9110は、ジャーナル作成処理を継続する。ジャーナル作成処理は、初期コピー処理内の処理であるか、命令受信処理内の処理であるかによって動作が異なる(ステップ1320)。ジャーナル作成処理が命令受信処理内の処理の場合は、ステップ130からの動作が行われる。ジャーナル作成処理が初期コピー処理内の場合は、ステップ1370からの動作が行われる。

[0099]

(3) ジャーナル作成処理が命令受信処理内の処理の場合、チャネルアダプタ110は、ライト対象の論理アドレス「A」が、初期コピー処理の処理対象となったかを調べる(ステップ1330)。論理ボリューム「A」のペア状態が"コピー未"の場合は、後に初期コピー処理にでジャーナル作成処理が行われるため、チャネルアダプタ110は、ジャーナルを作成せずに処理を終了する(ステップ1335)。論理ボリューム「A」のペア状態が"コピー中"の場合は、コピー済みアドレスが論理アドレス内位置「A」と等しいか小さいならば、後に初期コピー処理にてジャーナル作成処理が行われるため、チャネルアダプタ110は、ジャーナルを作成せずに処理を終了する(ステップ1335)。上記以外、つまり、論理ボリューム「A」のペア状態が"コピー中"かつコピー済みアドレスが論理アドレス内位置「A」以上の場合もしくは、論理ボリューム「A」のペア状態が

"正常"の場合は、既に初期コピー処理が完了しているため、チャネルアダプタ110は、ジャーナル作成処理を継続する。

[0100]

(4) 次に、チャネルアダプタ110は、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。すなわち、チャネルアダプタ110は、ポインタ情報700を用い、更新情報領域の未使用領域の有無を調べる(ステップ1340)。ポインタ情報700の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ110は、ジャーナル作成失敗として処理を終了する(ステップ1390)。

[0101]

ステップ1340の調べで、更新情報領域に未使用領域が存在する場合は、チャネルアダプタ110は、ポインタ情報700を用い、ライトデータ領域にライトデータが格納できるかを調べる(ステップ1345)。ライトデータ最新アドレスとデータ量「A」の和が、ライトデータ最古アドレスと等しいか大きい場合は、ライトデータ領域に格納できないため、チャネルアダプタ110は、ジャーナル作成失敗として処理を終了する(ステップ1390)。次のセットのポインタ情報700が存在する限り、チャネルアダプタ110は、次セットのポインタ情報700について、ステップ1340~ステップ1345の処理を繰り返す。

[0102]

ジャーナルが格納可能である場合、チャネルアダプタ110は、更新番号と、 更新情報を格納するための論理アドレスと、ライトデータを格納するための論理アドレス を取得し、更新情報をキャッシュメモリ130内に作成する。チャネルアダプタ110は 、対象グループのグループ情報600から更新番号を取得し、そして、これに1を加算し た数値を、グループ情報600の更新番号に設定する。チャネルアダプタ110は、更新 情報を格納するための論理アドレスとして、ポインタ情報700の更新情報最新アドレス を取得し、そして、それに更新情報のサイズを加算した数値を、ポインタ情報700の更 新情報最新アドレスに設定する。チャネルアダプタ110は、ライトデータを格納するた めの論理アドレスとして、ポインタ情報700のライトデータ最新アドレスを取得し、そ して、このライトデータ最新アドレスにデータ量「A」を加算した数値を、ポインタ情報 700のライトデータ最新アドレスに設定する。チャネルアダプタ110は、上記取得し た数値と、グループ番号と、ライト命令を受信した時刻と、ライト命令内の論理アドレス 「A」及びデータ量「A」を、更新情報に設定する(ステップ1350、図12の112 0)。例えば、図7に例示したグループ情報600及び図8に例示したポインタ情報70 0が存在する場合において、グループ「1」に属する正論理ボリューム「1」の記憶領域 の先頭からアドレス数「800」だけ進んだ位置にデータサイズ「100」のデータをラ イトせよというライト命令を受信した場合、チャネルアダプタ110は、図15に例示す るような更新情報を作成することになる。そして、図7に例示したグループ情報600内 の更新番号は「5」に変更される。また、ポインタ情報600内の更新情報最新アドレス は「600」(更新情報のサイズを「100」とした場合)に変更され、ライトデータ最 新アドレスは「2300」に変更される。正記憶システムに接続された複数の副記憶シス テムにそれぞれ対応する複数セットのポインタ情報700が存在する場合、チャネルアダ プタ110は、それら複数セットのポインタ情報700の全てを上記のようにして更新す

[0103]

(6) チャネルアダプタ110は、ディスクアダプタ120に、ジャーナルの更新情報とライトデータを記憶装置150に書き込むことを命令し、正常終了する(ステップ1360、図12の1130、1140、1150)。

$[0\ 1\ 0\ 4]$

(7) ジャーナル作成処理が、初期コピー処理内の処理の場合は、ステップ1370 からの動作が行われる。チャネルアダプタ110は、ジャーナルが作成可能であるかを調 べる。すなわち、チャネルアダプタ110は、ポインタ情報700を用い、更新情報領域の未使用領域の有無を調べる(ステップ1370)。ポインタ情報700の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ110は、ジャーナル作成失敗として処理を終了する(ステップ1390)。ここで、本実施形態における初期コピー処理においては、ジャーナルのライトデータは、正論理ボリュームからリードされ、ライトデータ領域は使用されないため、ライトデータ領域の未使用領域の確認は不要である。次のセットのポインタ情報700が存在する限り、チャネルアダプタ110は、次セットのポインタ情報700についてステップ1370の処理を繰り返す。

[0105]

(8) ステップ1370の調べで、ジャーナルが作成可能である場合、チャネルアダプタ110は、更新情報に設定すべき以下のような数値を取得し、更新情報をキャッシュメモリ130内に作成する。すなわち、チャネルアダプタ110は、対象グループのグループ情報600から更新番号を取得し、そして、これに1を足した数値をグループ情報600の更新番号に設定する。チャネルアダプタ110は、更新情報を格納するための論理アドレスとして、ポインタ情報700内の更新情報最新アドレスの位置を取得し、そして、これに更新情報のサイズを足した数値をポインタ情報700内の更新情報最新アドレスに設定する。

[0106]

チャネルアダプタ110は、上記取得した数値と、グループ番号と、本処理の開始時刻と、初期コピー処理対象の論理アドレスと、初期コピーの1回の処理データ量と、初期コピー処理におけるライトデータを格納したジャーナル論理ボリュームの論理アドレスとを、更新情報に設定する(ステップ1380、図12の1120)。

[0107]

(9) チャネルアダプタ 1 1 0 は、ディスクアダプタ 1 2 0 に、更新情報を物理的記憶装置 1 5 0 に書き込むことを命令し、正常終了する(ステップ 1 3 8 5 、図 1 2 0 1 1 4 0 、 1 1 5 0)。

[0108]

上記説明では、更新情報はキャッシュメモリ130内に格納されるが、変形として、更 新情報が共有メモリ140又はその他の記憶場所に格納されてもよい。

[0109]

ディスクアダプタ120によるライトデータの物理的記憶装置150への書き込みは、非同期で行うことができる(つまり、ステップ1360およびステップ1385の直後でなくともよい)。ただし、ホストコンピュータ180が、論理アドレス「A」に対するライト命令を再び発行した場合には、ジャーナルのライトデータが上書きされることになる。そのため、ホストコンピュータ180からライトデータを受信する前に、ジャーナルのライトデータは、更新情報の示すジャーナル論理ボリュームの論理アドレスに対応する物理記憶装置150内の記憶場所に、書き込まれる必要がある。もしくは、ジャーナルのライトデータは、別のキャッシュメモリに退避されて、後に更新情報のジャーナル論理ボリュームの論理アドレスに対応する物理的記憶装置150内の記憶場所に書き込まれてもよい。

$[0\ 1\ 1\ 0\]$

前述したジャーナル作成処理では、ジャーナルは適当なタイミングでキャッシュメモリから物理的記憶装置 150に保存される。変形例として、ジャーナル用に予め一定量のキャッシュメモリ130を用意しておき、当該キャッシュメモリが全て使用された段階で、ジャーナルが物理的記憶装置 150に保存されるようにしてもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定できるようにしてよい。

[0111]

リードライト処理220 (図2、図12参照)は、ディスクアダプタ120が、チャネルアダプタ110もしくはディスクアダプタ120から命令を受け、実施する処理である

。リードライト処理220は、指定されたキャッシュメモリ130のデータを、指定された論理アドレスに対応する物理的記憶装置150内の記憶領域に書き込むライト処理、及び、指定された論理アドレスに対応する記憶装置150内の記憶領域のデータを、指定されたキャッシュメモリ130に読み出すリード処理等に分類される。

[0112]

図16は、ジャーナルリード命令を受信した正記憶システム100Aのチャネルアダプタ110の動作(ジャーナルリード受信処理)におけるデータ流れを説明する図です。図17は、ジャーナルリード受信処理の手順を示すフローチャートである。以下、これらの図面を用いて、正記憶システム100Aが、副記憶システム100Bからジャーナルリード命令を受信した場合の動作について説明する。

[0113]

(1) 図16及び図17に示すように、正記憶システム100A内のチャネルアダプタ110は、副記憶システム100Bからアクセス命令(ジャーナルリード命令)を受信する。このアクセス命令は、ジャーナルリード命令であることを示す識別子、命令対象のグループ番号、及びリトライ指示の有無などを゛含んでいる(ステップ1220、図16の1410)。以下の説明では、そのアクセス命令(ジャーナルリード命令)内のグループ番号をグループ番号「A|とする。

[0114]

[0115]

(3) ステップ1510の調べで、グループ番号「A」のグループ状態が"正常"の場合、チャネルアダプ9110は、ジャーナル論理ボリュームの状態を調べる(ステップ1520)。ステップ1520の調べで、ジャーナル論理ボリュームのボリューム状態が"正常"でない場合、例えば、"障害"の場合は、チャネルアダプ9110は、グループ状態を"障害"に変更し、副記憶システム100Bは、受信したグループ状態に応じて処理を行う。例えば、副記憶システム100Bは、グループ状態が"障害"の場合は、ジャーナルリード処理を終了する(ステップ1525)。

[0116]

(4) ステップ1520の調べで、ジャーナル論理ボリュームのボリューム状態が "正常"の場合は、チャネルアダプタ110は、ジャーナルリード命令がリトライ指示かを調べる(ステップ1530)。

[0117]

(5) ステップ1530の調べで、ジャーナルリード命令がリトライ指示の場合、チャネルアダプ9110は、前回送信したジャーナルを再度、副記憶システム100Bに送信する。チャネルアダプ91100は、キャッシュメモリ1300を確保し、ディスクアダプ9120に、ポインタ情報7000のリトライ開始アドレスが指す記憶場所から、更新情報のサイズ分のデータ(更新情報)をキャッシュメモリに読み込むことを命令する(図1601420)。

[0118]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150から更新情報を読み込み、これをキャッシュメモリ130に保存し、更新情報のリード終了をチャネルアダプタ110に通知する(図16の1430)。

[0119]

チャネルアダプタ110は、更新情報のリード終了の通知を受け、更新情報から、ライ

トデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ130を確保し、ディスクアダプタ120にライトデータをキャッシュメモリ130に読み込むことを命令する(ステップ1540、図16の1440)。

[0120]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150からライトデータを読み込み、これをキャッシュメモリ130に保存し、ライトデータのリード終了をチャネルアダプタ110に通知する(図16の1450)。

[0121]

チャネルアダプタ110は、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム100Bに送信し、ジャーナルを保持しているキャッシュメモリ130を開放し、処理を終了する(ステップ1545、図16の1460)。

[0122]

(6) ステップ1530の調べで、リトライ指示でない場合、チャネルアダプタ110は、送信していないジャーナルが存在するかを調べ、存在すれば、そのジャーナルを副記憶システム100Bに送信する。チャネルアダプタ110は、ポインタ情報700のリード開始アドレスと更新情報最新アドレスを比較する(ステップ1550)。

[0123]

リード開始アドレスが更新情報最新アドレスと等しい場合は、全てのジャーナルを副記憶システム100Bに送信済みであるため、チャネルアダプタ110は、副記憶システム100Bに"ジャーナル無"を送信し(ステップ1560)、そして、前回のジャーナルリード命令の処理時に副記憶システム100Bに送信したジャーナルの記憶領域を開放する(ステップ1590)。

[0124]

ステップ1590のジャーナルの記憶領域の開放処理では、その記憶領域に対して積極的にこれを開放する動作を行う必要はなく、単に最古アドレスのポインタを更新するだけでよい。すなわち、ポインタ情報700の更新情報最古アドレスに、リトライ開始アドレスが設定される。更新情報最古アドレスがライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは"0"とされる。ポインタ情報700のライトデータ最古アドレスは、その現在値に、前回のリードジャーナル命令に応じて送信されたライトデータのサイズを加算した数値に変更される。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスが減じられるよう補正される。

[0125]

(7) ステップ1550の調べで、未送信のジャーナルが存在する場合、チャネルアダプタ110は、キャッシュメモリ130を確保し、ディスクアダプタにポインタ情報700のリード開始アドレスから、更新情報のサイズ分のデータ(更新情報)をキャッシュメモリに読み込むことを命令する(図16の1420)。

[0126]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150から更新情報を読み込み、これをキャッシュメモリ130に保存し、更新情報のリード終了をチャネルアダプタ110に通知する(図16の1430)。

$[0\ 1\ 2\ 7\]$

チャネルアダプタ110は、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ130を確保し、ディスクアダプタ120にライトデータをキャッシュメモリに読み込むことを命令する(ステップ1570、図16の1440)。

[0128]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150からライトデータを読み込み、これをキャッシュメモリ130に保存し、ライトデータのリード終了をチャネルアダプタに通知する(図16の1450)。

[0129]

チャネルアダプタ110は、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム100Bに送信し(ステップ1580)、ジャーナルを保持しているキャッシュメモリ130を開放する(図16の1460)。そして、ポインタ情報700のリトライ開始アドレスにリード開始アドレスを設定し、そして、リード開始アドレスを、その現在値に、送信されたジャーナルの更新情報サイズを加算した数値に変更する。

[0130]

(8) チャネルアダプタ110は、前回のジャーナルリード命令の処理時に副記憶システム100Bに送信されたジャーナルの記憶領域を開放する(ステップ1590)。

[0131]

上述したジャーナルリード受信処理では、正記憶システム100Aは、ジャーナルを一つずつ逐次に副記憶システム100Bに送信する。変形例として、複数のジャーナルを同時に副記憶システム100Bに送信するようにしてもよい。1つのジャーナルリード命令で送信されるジャーナル数は、副記憶システム100Bによりジャーナルリード命令内で指定されてもよいし、或いは、グループ登録の時等に、ユーザにより正記憶システム100A又は副記憶システム100Bに指定されてもよい。さらに、正記憶システム100Aと副記憶システム100Bの接続パス200の転送能力又は負荷等に応じて動的に、1つのジャーナルリード命令で送信されるジャーナル数が変更されてもよい。また、ジャーナル数でなく、ジャーナルのライトデータのサイズに応じて、ジャーナルの転送量が指定されてもよい。

[0132]

前述したジャーナルリード受信処理では、ジャーナルが物理的記憶装置150からキャッシュメモリ130に読み込まれる。しかし、この動作は、ジャーナルが既にキャッシュメモリ130に存在する場合は不要である。

[0 1 3 3]

前述したジャーナルリード受信処理内のジャーナルの記憶領域の開放処理は、次のジャーナルリード命令の処理時に開放される。変形例として、副記憶システム100Bにジャーナルを送信した直後に、そのジャーナルの記憶領域が開放されてもよい。また、副記憶システム100Bが、ジャーナルリード命令内で開放してよい更新番号を設定し、正記憶システム100Aは、その設定により指定された更新番号のジャーナルの記憶領域を開放するようにしてもよい。

$[0\ 1\ 3\ 4]$

図18は、ジャーナルリード(JNLRD)処理240におけるデータ流れを説明する図である。図19は、ジャーナルリード処理240の手順を示すフローチャートである。図20は、ジャーナルリード処理240の中で行われるジャーナル格納処理の手順を示すフローチャートである。以下、これらの図面を用いて、副記憶システム100Bのチャネルアダプタ110が、正記憶システム100Aからジャーナルを読み出し、ジャーナル論理ボリュームに格納する動作について説明する。

[0135]

(1) 副記憶システム 100 B内のチャネルアダプタ 110 は、ジャーナルを格納するキャッシュメモリ 130 を確保し、ジャーナルリード命令であることを示す識別子、命令対象の正記憶システム 100 Aのグループ番号、及びリトライ指示の有無を含むアクセス命令(ジャーナルリード命令)を、正記憶システム 100 Aに送信する(ステップ 170 Colon (2000) の 180 Colon (2000

[0136]

(2) チャネルアダプタ110は、正記憶システム100Aの応答およびジャーナルを受信する(図18の1620)。チャネルアダプタ110は、正記憶システム100Aからの応答を調べ、その応答が、"ジャーナル無"の場合は、正記憶システム100Aに

は、指定したグループのジャーナルが存在しないため、一定時間後、正記憶システム 10 0Aにリードジャーナル命令を送信する(ステップ 1720、1725)。

[0137]

(4) 正記憶システム 100 A の応答が、"グループ状態は障害"もしくは"グループ状態は未使用"の場合は、チャネルアダプタ 110 は、副記憶システム 100 B のグループ状態を受信した状態に変更し、ジャーナルリード処理を終了する(ステップ 1730 、 1735)。

[0138]

(5) 正記憶システム100Aの応答が、上記以外、つまり、正常終了の場合は、チャネルアダプタ110は、ジャーナル論理ボリュームのボリューム状態を調べる(ステップ1740)。ジャーナル論理ボリュームのボリューム状態が "異常"の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、チャネルアダプタ110は、グループ状態を "異常"に変更し、処理を終了する(ステップ1745)。この場合、チャネルアダプタ110は、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行い、グループの状態を正常に戻す。

[0139]

(6) ステップ1740の調べで、ジャーナル論理ボリュームのボリューム状態が"正常"の場合は、チャネルアダプタ110は、後述するジャーナル格納処理1800を行う。ジャーナル格納処理1800が正常に終了した場合は、チャネルアダプタ110は、直ちに次のジャーナルリード命令を送信するか、もしくは、一定時間経過後、次のジャーナルリード命令を送信する(ステップ1760)。次のジャーナル命令を送信するタイミングは、一定の時間間隔で定期的に送信するものでもよく、その時間間隔は、受信したジャーナルの個数、接続線200の通信量、副記憶システム100Bが保持しているジャーナルの記憶容量、又は副記憶システム100Bの負荷等に応じて制御されてもよい。さらに、正記憶システム100Aが保持しているジャーナルの記憶容量又は正記憶システム100Aのポインタ情報を、副記憶システム100Bが読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情報の転送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれていてよい。その後の処理は、ステップ1710以降と同じである。

$[0\ 1\ 4\ 0\]$

(7) ステップ1800のジャーナル格納処理が正常に終了しない場合は、ジャーナル論理ボリュームの未使用領域が足りないため、チャネルアダプタ110は、受信したジャーナルを破棄し、一定時間後にリトライ指示のジャーナルリード命令を送信する(ステップ1755)。もしくは、チャネルアダプタ110は、ジャーナルをキャッシュメモリに保持しておき、一定時間後に、再度ジャーナル格納処理を行う。これは、後述するリストア処理250が行われることにより、一定時間後には、ジャーナル論理ボリュームに未使用領域が増える可能性があるためである。この方式の場合は、ジャーナルリード命令にリトライ指示の有無は不要である。

[0141]

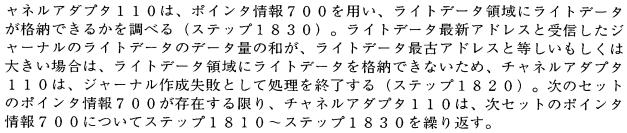
次に、上述したジャーナル格納処理1800について、図20を参照して説明する。

[0142]

(1) 副記憶システムチャネルアダプタ110は、最初のセットのポインタ情報700を取得する(ステップ1805)。副記憶システムチャネルアダプタ110は、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。すなわち、チャネルアダプタ110は、ポインタ情報700を用い、更新情報領域に未使用領域の有無を調べる(ステップ1810)。ポインタ情報700の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ110は、ジャーナル作成失敗として処理を終了する(ステップ1820)。

[0 1 4 3]

(2) ステップ1810の調べで、更新情報領域に未使用領域が存在する場合は、チ



[0144]

(3) ジャーナルが格納可能である場合、チャネルアダプタ110は、受信した更新情報のグループ番号とジャーナル論理ボリュームの論理アドレスを変更する。グループ番号は、副記憶システム100Bのグループ番号に変更され、ジャーナル論理ボリュームの論理アドレスはポインタ情報700のライトデータ最新アドレスに変更される。チャネルアダプタ110は、ポインタ情報700の更新情報最新アドレスを、これに更新情報最新アドレスに更新情報のサイズを足した数値に変更する。チャネルアダプタ110は、ポインタ情報700のライトデータ最新アドレスを、このライトデータ最新アドレスにライトデータのサイズを足した数値に変更する(ステップ1840)。複数の副記憶システムにそれぞれ対応する複数セットのポインタ情報が存在する場合、チャネルアダプタ110は、それら複数セットのポインタ情報700の全てを上記のように更新する。

[0145]

(4) チャネルアダプタ110は、ディスクアダプタ120に、更新情報とライトデータを記憶装置150に書き込むことを命令し、ジャーナル作成成功として処理を終了する(ステップ1850、図18の1630)。その後、ディスクアダプタ120は、リードライト処理220により、物理的記憶装置150に更新情報とライトデータを書き込み、キャッシュメモリ130を開放する(図18の1640)。

[0146]

上述したジャーナル格納処理では、ジャーナルが適当なタイミングでキャッシュメモリ 130から物理的記憶装置 150に保存される。変形例として、ジャーナル用に予め一定量のキャッシュメモリ 130を用意しておき、当該キャッシュメモリが全て使用された段階で、物理的記憶装置 150にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定できるようにしてよい。

$[0\ 1\ 4\ 7\]$

図21は、リストア処理250におけるデータ流れを説明する図である。図22は、リストア処理250の手順を示すフローチャートである。以下、これらの図面を用いて、副記憶システム100Bのチャネルアダプタ110が、ジャーナルを利用し、データの更新を行う動作について説明する。変形例として、リストア処理250が、副記憶システム100Bのディスクアダプタ120により行われてもよい。

[0148]

(1) 副記憶システム100Bチャネルアダプタ110は、グループ番号Bのグループ状態が"正常"であるかを調べる(ステップ2010)。ステップ2010の調べで、グループ状態が"正常"以外、例えば、"障害"の場合は、チャネルアダプタ110は、リストア処理を終了する(ステップ2015)。

[0149]

(2) ステップ2010の調べで、グループ状態が"正常"の場合は、チャネルアダプタ110は、ジャーナル論理ボリュームのボリューム状態を調べる(ステップ2020)。ステップ2020の調べで、ジャーナル論理ボリュームのボリューム状態が、"異常"の場合は、アクセス不可能なため、チャネルアダプタ110は、グループ状態を"異常"に変更し、処理を終了する(ステップ2025)。

[0150]

(3) ステップ2020の調べで、ジャーナル論理ボリュームのボリューム状態が、 "正常"の場合は、チャネルアダプタ110は、リストア対象のジャーナルが存在するか

を調べる。すなわち、チャネルアダプタ110は、ポインタ情報700の更新情報最古アドレスと更新情報最新アドレスを取得する。更新情報最古アドレスと更新情報最新アドレスが等しい場合、ジャーナルは存在しないため、チャネルアダプタ110は、リストア処理は一旦終了し、一定時間後、リストア処理を再開する(ステップ2030)。

[0151]

(4) ステップ2030の調べで、リストア対象のジャーナルが存在する場合、チャネルアダプタ110は、最古(最小)の更新番号を持つジャーナルに対して次の処理を行う。最古(最小)の更新番号を持つジャーナルの更新情報は、ポインタ情報700の更新情報最古アドレスから保存されている。チャネルアダプタ110は、キャッシュメモリ130を確保し、ディスクアダプタ120に、更新情報最古アドレスから、更新情報のサイズ分のデータ(更新情報)をキャッシュメモリ130に読み込むことを命令する(図21の1910)。

[0152]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150から更新情報を読み込み、キャッシュメモリ130に保存し、更新情報のリード終了をチャネルアダプタ110に通知する(図21の1920)。

[0153]

チャネルアダプタ110は、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ130を確保し、ディスクアダプタ120にライトデータをキャッシュメモリに読み込むことを命令する(図21の1930)。

[0154]

ディスクアダプタ120のリードライト処理220は、物理的記憶装置150からライトデータを読み込み、キャッシュメモリ130に保存し、ライトデータのリード終了をチャネルアダプタ110に通知する(ステップ2040、図21の1940)。

[0155]

(5) チャネルアダプタ110は、更新情報から更新する副論理ボリュームの論理アドレスを求め、ディスクアダプタ120に副論理ボリュームにライトデータを書き込むことを命令する(ステップ2050、図21の1950)。ディスクアダプタ120のリードライト処理220は、副論理ボリュームの論理アドレスに対応する物理的記憶装置150の記憶場所にデータを書き込み、キャッシュメモリ130を開放し、データのライト完了をチャネルアダプタ110に通知する(図21の1960)。

[0156]

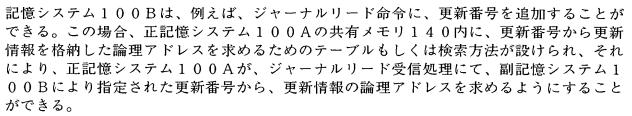
(6) チャネルアダプタ110は、データのライト完了の通知を受け、ジャーナルの記憶領域を開放する。ジャーナルの記憶領域の開放処理では、ポインタ情報700の更新情報最古アドレスが、これに更新情報のサイズを足した数値に変更される。更新情報最古アドレスが、ライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは"0"とされる。ポインタ情報700のライトデータ最古アドレスは、これにライトデータのサイズを足した数値に変更される。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスが減じされるよう補正される。その後、チャネルアダプタ110は、次のリストア処理を開始する(ステップ2060)。

[0157]

前述したリストア処理250では、物理的記憶装置150からキャッシュメモリ130 にジャーナルが読み込まれる。しかし、この動作は、ジャーナルが既にキャッシュメモリ 130に存在する場合は不要である。

[0158]

前述したジャーナルリード受信処理とジャーナルリード処理240では、正記憶システム100Aが、送信すべきジャーナルをポインタ情報700により決めている。変形例として、副記憶システム100Bが、送信すべきジャーナルを決めてもよい。この場合、副



[0159]

前述したジャーナルリード受信処理とジャーナルリード処理240では、ジャーナルリード命令という専用のアクセス命令が用いられる。変形例として、通常のリード命令を用いるようにしてもよい。その場合、例えば、正記憶システム100Aのグループ情報600とポインタ情報700が予め副記憶システム100Bに転送され、そして、副記憶システム100Bが、グループ情報600とポインタ情報700に基づいて、正記憶システム100A内のジャーナル論理ボリュームのデータ(つまり、ジャーナル)をリードするためのリード命令を生成するようにすることができる。

[0160]

前述したジャーナルリード受信処理では、更新番号の順に、正記憶システム100Aから副記憶システム100Bにジャーナルが順次に送信される。変形例として、更新番号の順とは異なる順序でジャーナルが送信されるようにしてもよい。或いは、正記憶システム100Aから副記憶システム100Bに、複数のジャーナルを並列的に送信されてもよい。この場合、副記憶システム100Bに、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法が設けられ、それにより、副記憶システム100Bでのリストア処理250で、更新番号順にジャーナルが処理されるようにすることができる

[0161]

本実施形態では、正記憶システム100Aがジャーナルを取得し、副記憶システム100Bが、正記憶システム100Aからジャーナルをリードして、どれに基づいてデータの複製を行う。これにより、正記憶システム100Aに接続されたホストコンピュータ180は、データの複製に関する負荷を負わない。さらに、正記憶システム100Aと副記憶システム100B間でジャーナルが転送されるので、正記憶システム100Aとホストコンピュータ180間の通信線がデータ複製のためには使用されない。

[0162]

図23は、本発明に従うデータ処理システムの第2の実施形態の論理的な構成を示す図である。

[0163]

図23に示すように、この実施形態は、正記憶システム100Aと副記憶システム100Bの他に、第3の記憶システム100Cを有する。これらの記憶システム100A、100B、100Cの物理的な構成は、いずれも、図1を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ180と第3の記憶システム100Cが接続パス190により接続され、第3の記憶システム100Cと正記憶システム100Aが接続パス200により接続され、そして、正記憶システム100Aと副記憶システム100Bが接続パス200により接続される。第3の記憶システム100Cは、正記憶システム100A内の正論理ボリューム(「DATA1」、「DATA2」等)230内のデータの元のデータをそれぞれ保持したオリジナル論理ボリューム(「ORG1」、「ORG2」等)230を有する。

$[0\ 1\ 6\ 4]$

第3の記憶システム100Cは、ホストコンピュータ180からのデータライト命令により、要求されたオリジナル論理ボリューム(例えば「ORG1」)230内のデータ(オリジナルデータ)を更新する。この時、第3の記憶システム100Cは、オリジナル論理ボリューム(例えば「ORG1」)230内のオリジナルデータを更新するだけでなく、その更新対象のオリジナルデータに対応する正論理ボリューム(「DATA1」)23



0内のデータを更新するためのデータライト命令を、正記憶システム 1 0 0 A に送る (2 3 1 0)。

[0165]

正記憶システム 100 A は、第 1 の実施形態で説明した通り、上記データライト命令を受けて、要求された正論理ボリューム(例えば「DATA1」) 230 のデータを更新し、そして、前述した命令受信処理 210 およびリードライト処理 220 によって、そのデータ更新のジャーナルを、ジャーナル論理ボリューム(「JNL1」) 230 に保存する(2310)。

[0166]

副記憶システム100Bは、前述したジャーナルリード処理240によって、正記憶システム100Aからジャーナルをリードし、リードライト処理220によって、ジャーナル論理ボリューム(「JNL2」)230にジャーナルを保存する(2320)。

[0167]

正記憶システム 100 A は、副記憶システム 100 B からジャーナルをリードする命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム(「JNL1」) 230 からジャーナルを読み出し、副記憶システム 100 B に送信する (2320)。

[0168]

副記憶システム100Bは、前述したリストア処理250およびリードライト処理22 0によって、更新番号に従い、ジャーナル論理ボリューム(「JNL2」)からジャーナルを読み出し、正論理ボリューム(「DATA1」)230の複製である副論理ボリューム(COPY1)のデータを更新する(2330)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

$[0 \ 1 \ 6 \ 9]$

図23に示したデータ処理システムでは、正記憶システム100Aがジャーナルを取得し、これをジャーナル専用の記憶領域に格納する。さらに、副記憶システム100Bは、正記憶システムから受信したジャーナルを、ジャーナル専用の記憶領域に格納する。ジャーナル専用の記憶領域は、データ複製対象の記憶領域より少なくすることが可能であり、より少ない記憶容量で、副記憶システムに正記憶システムのデータの複製が可能となる。

[0170]

図24は、本発明に従うデータ処理システムの第3の実施形態の論理的な構成を示す図である。

[0171]

図24に示すように、この実施形態は、正記憶システム100Aと副記憶システム100Bの他に、第3の記憶システム100Cを有する。これらの記憶システム100A、100B、100Cの物理的な構成は、いずれも、図1を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ180と第3の記憶システム100Cが接続パス190により接続され、第3の記憶システム100Cと正記憶システム100Aが接続パス200により接続され、そして、正記憶システム100Aと副記憶システム100Bが接続パス200により接続される。

[0172]

正記憶システム100Aは、第3の記憶システム100Cに対し、正論理ボリューム(「DATA1」、「DATA2」等)があるように見せるが、その正論理ボリューム(「DATA1」、「DATA2」等)に対しては、実際の物理的な記憶領域、つまり物理的記憶装置150を割り当てない。例えば、ボリューム情報400内の各正論理ボリュームの物理アドレスに、物理的記憶装置150を割り当てていないことを示す所定の数値が設定される。従って、正記憶システム100A内の正論理ボリューム(「DATA1」、「DATA2」等)は、仮想的なものである。正記憶システム100Aは、これら仮想的な正論理ボリューム(「DATA1」、「DATA2」等)のデータ更新のジャーナルを保持するためのジャーナル論理ボリューム(「JNL1」)230を有し、これには、実際



の物理的な記憶領域が割り当てられる。第3の記憶システム100Cは、正記憶システム100A内の仮想的な正論理ボリューム(「DATA1」、「DATA2」)内のデータに相当する実際のデータを保持したオリジナル論理ボリューム(例えば「ORG1」、「ORG2」等)230を有し、これには実際の物理的記憶領域が割り当てられている。

[0173]

第3の記憶システム100Cは、ホストコンピュータ180からのデータライト命令により、要求されたオリジナル論理ボリューム(例えば「ORG1」)のデータ(オリジナルデータ)を更新する。この時、第3の記憶システム100Cは、そのオリジナルデータを更新するだけでなく、更新されるオリジナルデータに対応する仮想的な正論理ボリューム(例えば「DATA1」)内のデータを更新するためのデータライト命令を、正記憶システム100Aに送る(2410)。

[0174]

正記憶システム 100 A は、第3の記憶システム 100 C から仮想的な正論理ボリューム(例えば「DATA1」)内のデータのライト命令を受信すると、図 13 に示した命令受信処理 210 のステップ 1270 の処理(ディスクアダプタへのデータライト命令の発行)は行わずに、そのデータ更新のジャーナルをジャーナル論理ボリューム(「JNL1」) 230 に保存するだけである (2410)。

[0175]

副記憶システム 100B は、前述したジャーナルリード処理 240 によって、正記憶システム 100A からジャーナルをリードし、リードライト処理 220 によって、ジャーナル論理ボリューム(「JNL2」) 230 にジャーナルを保存する(2420)。

[0176]

正記憶システム 100 A は、副記憶システム 100 B からジャーナルリード命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム(「JNL1」) 230 からジャーナルを読み出し、記憶システム 100 B に送信する (2420)。

[0177]

副記憶システム100Bは、前述したリストア処理250およびリードライト処理220によって、更新番号に従い、ジャーナル論理ボリューム(「JNL2」)230からジャーナルを読み出し、オリジナル論理ボリューム(例えば「ORG1」)230の複製である副論理ボリューム(例えば「COPY1」)230のデータを更新する(2430)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

[0178]

図24に示したデータ処理システムでは、第3の記憶システム100Cもしくは、第3の記憶システム100Cに接続されたホストコンピュータ180に障害が生じた場合、副記憶システム100Bの論理ボリューム(例えば「COPY1」)230に対し、正記憶システム100A内のジャーナル(「JNL1」)230を反映することにより、副記憶システム100Bに接続されたホストコンピュータ(図24には図示せず)により、最新データの参照、更新が可能となる。さらに、正記憶システム100Aにオリジナルデータの複製を保持せず、ジャーナルのみを格納することで、データ複製に必要な記憶容量が少なくすることが可能となる。

[0179]

図25は、本発明に従うデータ処理システムの第4の実施形態の論理的な構成を示す図である。

[0180]

図25に示すように、この実施形態は、正記憶システム100Aと複数(例えば2台)の副記憶システム100B、100Cを有する。これらの記憶システム100A、100B、100Cの物理的な構成は、いずれも、図1を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ180と正記憶システム100Aが接続パス190によ



り接続され、正記憶システム100Aと第1の副記憶システム100Bが接続パス200により接続され、そして、正記憶システム100Aと第2の副記憶システム100Cが接続パス200により接続される。正記憶システム100Aでは、所定の複数の正論理ボリューム(例えば「DATA1」、「DATA2」)230とジャーナル論理ボリューム(例えば「JNL1」)230が一つのグループ「グループ1」を構成する。第1の副記憶システム100Bでは、上述した「グループ1」に属する複数の正論理ボリューム(「DATA1」、「DATA2」)230のそれぞれの複製である複数の副論理ボリューム(例えば「COPY1」、「COPY2」)230とジャーナル論理ボリューム(例えば「JNL2」)230が同じグループ「グループ1」を構成する。同様に、第2の副記憶システム100Cでも、「グループ1」の正論理ボリューム(「DATA1」、「DATA2」)230のそれぞれの複製である複数の副論理ボリューム(例えば「JNL3」)230が同じグループ1」の正論理ボリューム(例えば「JNL3」)230が同じグループ「グループ1」を構成する。

[0181]

1台の正記憶システム100Aに対して複数台の副記憶システム100B、100Cが存在するため、図26に例示するように、複数台の副記憶システム100B、100Cにそれぞれ対応した複数のポインタ情報700B、700Cが、正記憶システム100Aに保持される。第1の副記憶システム100Bには、第1の副記憶システム100B用のポインタ情報700Bが保持され、第2の副記憶システム700Cにも、第2の副記憶システム700Cにも、第2の副記憶システム700Cの番々の構成とその意味するところは、図8及び図9を参照して既に説明したものと同様である。異なる副記憶システム100B、100Cは、それぞれ独自にスケジューリングしたタイミングで、ジャーナルリード処理240及びリストア処理250を行う。そのため、図27に例示するように、異なる副記憶システム100B、100C用のポインタ情報700B、700Cが指し示すアドレスは必ずしも同じではない。

[0182]

再び図25を参照して、この実施形態の動作を以下に説明する。

[0183]

正記憶システム 100 A は、ホストコンピュータ 180 から或る正論理ボリューム(例えば「DATA1」) 230 のデータのライト命令を受信すると、前述した命令受信処理 210 およびリードライト処理 220 によって、正論理ボリューム(「DATA1」) 230 内の要求されたデータを更新し、そして、ジャーナル論理ボリューム(「JNL1」)にそのデータ更新のジャーナルを保存する (2510)。

$[0\ 1\ 8\ 4]$

第1の副記憶システム100Bは、前述したジャーナルリード処理240によって、正 記憶システム100Aからジャーナルをリードし、リードライト処理220によって、ジ ャーナル論理ボリューム (「JNL2」) にそのジャーナルを保存する(2520)。ここ で、第1の副記憶システム100Bがジャーナルをリードするタイミング(ジャーナルリ ード命令を正記憶システム100Aに送るタイミング)は、第1の副記憶システム100 Bのチャネルアダプタ110によって独自にスケジューリングされる。このジャーナルリ ードのタイミングじゃ図19のステップ1760に関して既に説明したように、例えば、 前のジャーナルのジャーナル格納処理1800の正常終了後に直ちにでもよいし、前のジ ャーナルのジャーナル格納処理1800の正常終了後一定時間経過後でもよいし、或いは 、一定の時間間隔で定期的にでもよい。定期的にジャーナルリード命令を送る場合には、 その時間間隔は、受信したジャーナルの個数、接続線200の通信量、第1の副記憶シス テム100Bが保持しているジャーナルの記憶容量、又は第1の副記憶システム100B の負荷等に応じて制御されてもよい。さらに、正記憶システム100Aが保持しているジ ャーナルの記憶容量又は正記憶システム100Aのポインタ情報を、第1の副記憶システ ム100Bが読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情 報の転送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれ



ていてよい。いずれにしても、ジャーナルをリードするタイミングが他の副記憶システム 100Cと同期する必要はない。

[0185]

正記憶システム 100 A は、副記憶システム 100 B からジャーナルをリードする命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム(「JNL1」)からジャーナルを読み出し、第1の副記憶システム 100 B に送信する (2520)。

[0186]

第1の副記憶システム100Bは、前述したリストア処理250およびリードライト処理220によって、更新番号に従い、ジャーナル論理ボリューム(「JNL2」)からジャーナルを読み出し、正論理ボリューム(「DATA1」)の複製である副論理ボリューム(「COPY1」)のデータを更新する(290)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

[0187]

第2の副記憶システム100Cは、前述したジャーナルリード処理240によって、正 記憶システム100Aからジャーナルをリードし、リードライト処理220によって、ジ ャーナル論理ボリューム(「JNL3丨)にジャーナルを保存する(2520)。ここで、 第2の副記憶システム100Cがジャーナルをリードするタイミング(ジャーナルリード 命令を正記憶システム100Aに送るタイミング)は、第2の副記憶システム100Cの チャネルアダプタ110によって独自にスケジューリングされる。このジャーナルリード のタイミングは、図19のステップ1760に関して既に説明したように、例えば前のジ ャーナルのジャーナル格納処理1800の正常終了後に直ちにでもよいし、前のジャーナ ルのジャーナル格納処理1800の正常終了後一定時間経過後でもよいし、或いは、一定 の時間間隔で定期的にでもよい。定期的にジャーナルリード命令を送る場合には、その時 間間隔は、受信したジャーナルの個数、接続線200の通信量、第2の副記憶システム1 0 0 ℃が保持しているジャーナルの記憶容量、又は第2の副記憶システム100℃の負荷 等に応じて制御されてもよい。さらに、正記憶システム100Aが保持しているジャーナ ルの記憶容量又は正記憶システム100Aのポインタ情報を、第2の副記憶システム10 0 C が読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情報の転 送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれていて よい。いずれにしても、ジャーナルをリードするタイミングは、他の副記憶システム10 0 Bと同期する必要はない。

[0188]

正記憶システム 100Aは、第2の副記憶システム 100Cからジャーナルをリードする命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム(「JNL1」)からジャーナルを読み出し、第2の記憶システム 100Cに送信する (2520)。

[0189]

第2の副記憶システム100Cは、前述したリストア処理250およびリードライト処理220によって、更新番号に従い、ジャーナル論理ボリューム(「JNL3」)からジャーナルを読み出し、正論理ボリューム(「DATA1」)の複製である副論理ボリューム(「COPY3」)のデータを更新する(290)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

[0190]

上述したように、異なる副記憶システム100B、100Cが、それぞれ独自にスケジューリングしたタイミングでジャーナルリード処理240及びリストア処理250を行う。副記憶システム100B、100Cは、それぞれ、リストア処理(又はジャーナルリード処理)が終わると、リストア処理(又はジャーナルリード処理)の終了した更新番号を含むリストア処理(又はジャーナルリード処理)終了の通知を、正記憶システム100Aに送る。正記憶システム100Aは、副記憶システム100B、100Cからのリストア



処理(又はジャーナルリード処理) 終了の通知に基づいて、どの副記憶システムシステム 100 B、100 Cがどの更新番号のリストア処理(又はジャーナルリード処理) を終了したかを示す情報を管理する。そして、正記憶システム 100 A は、その情報に基づいて、全ての副記憶システム 100 B、100 Cにてリストア処理(又はジャーナルリード処理) が終了した更新番号のジャーナルについて、ジャーナル論理ボリューム(「JNL 1」) 230 内の当該ジャーナルの記憶領域を開放する。副記憶システム 100 B、100 Cのいずれか一つでも未だリストア処理(又はジャーナルリード処理) が終了してない更新番号のジャーナルについては、正記憶システム 100 A は、そのジャーナルをジャーナル論理ボリューム(「JNL 1」) 230 内に維持し、そのジャーナルの記憶領域を開放しない。

[0191]

図25に例示したデータ処理システムによれば、複数台のうちの一部の副記憶システム で障害が発生しても、他の正常な副記憶システムにより正論理ボリュームの複製が維持さ れるので、安全性が高い。

[0192]

図25に例示したデータ処理システムでは、一台の正記憶システム100Aに対して2台の副記憶システム100B、100Cが存在する。変形例として、3台以上の副記憶システム100B、100Cが一台の正記憶システム100Aに対して設けられてもよい。

[0193]

図25に例示したデータ処理システムでは、複数台の副記憶システム100B、100Cが並列的に一台の正記憶システム100Aからジャーナルをリードする。変形例として、各副記憶システム100B、100Cが、正記憶システム100Aからジャーナルをリードする機能の他に、他の副記憶システムからジャーナルをリードする機能も兼ね備えて、正記憶システム100Aと他の副記憶システムのどれからジャーナルをリードするかを選択できるようにすることもできる。例えば、正記憶システム100Aの負荷が小さいときは、全ての副記憶システム100B、100Cが、正記憶システム100Aからジャーナルをリードするが、他方、正記憶システム100Aの負荷が大きいときには、第1の副記憶システム100Bが正記憶システム100Bからそのジャーナルをリードするというような制御を行うことができる。

[0194]

以上、本発明の幾つかの実施形態を説明した。これらの実施形態によれば、記憶システムの上位の計算機に影響を与えずに、或いは、記憶システムと計算機との間の通信にも影響を与えずに、複数の記憶システム間でデータ転送又はデータの複製をすることが可能である。

[0195]

さらに、或る実施形態によれば、複数の記憶システム内に保持するデータ格納領域を少なくすることができる。また、或る実施形態によれば、複数の記憶システムの上位の計算機の業務に大きい影響を与えることのなしに、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることができる。

[0196]

本発明は、上述した実施形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【図面の簡単な説明】

[0197]

- 【図1】本発明の第1の実施形態のデータ処理システムの物理的構成を示すブロック 図である。
- 【図2】第1実施形態のデータ処理システムの論理的構成を示すブロック図である。
- 【図3】ジャーナルの更新情報とライトデータの関係を説明する図である。
- 【図4】ジャーナルに含まれる更新情報の具体例を示す図である。



- 【図5】第1実施形態のボリューム情報の例を説明する図である。
- 【図6】第1実施形態のペア情報の例を説明する図である。
- 【図7】第1実施形態のグループ情報の例を説明する図である。
- 【図8】第1実施形態のポインタ情報の例を説明する図である。
- 【図9】ポインタ情報内の項目とジャーナル論理ボリュームとの関係を説明する図である。
- 【図10】第1実施形態のデータの複製を開始する手順を説明するフローチャートである。
- 【図11】第1実施形態の初期コピー処理を説明するフローチャートである。
- 【図12】第1実施形態の命令受信処理におけるデータ流れを説明する図である。
- 【図13】第1実施形態の命令受信処理の手順を示すフローチャートである。
- 【図14】第1実施形態のジャーナル作成処理の手順を示すフローチャートである。
- 【図15】第1実施形態のジャーナル作成処理時の更新情報の例を説明する図である
- 【図16】第1実施形態のジャーナルリード受信処理におけるデータ流れを説明する 図である。
- 【図17】第1実施形態のジャーナルリード受信処理の手順を示すフローチャートである。
- 【図18】第1実施形態のジャーナルリード命令処理におけるデータ流れを説明する 図である。
- 【図19】第1実施形態のジャーナルリード命令処理の手順を示すフローチャートである。
- 【図20】第1実施形態のジャーナル格納処理の手順を示すフローチャートである。
- 【図21】第1実施形態のリストア処理におけるデータ流れを説明する図である。
- 【図22】第1実施形態のリストア処理の手順を示すフローチャートである。
- 【図23】本発明の第2の実施形態の論理的な構成を示す図である。
- 【図24】本発明の第3の実施形態の論理的な構成を示す図である。
- 【図25】本発明の第4の実施形態の論理的な構成を示す図である。
- 【図26】第4実施形態のポインタ情報の例を説明する図である。
- 【図27】第4実施形態のポインタ情報とジャーナル論理ボリュームとの関係を説明する図である。

【符号の説明】

[0198]

- 100 記憶システム
- 110 チャネルアダプタ
- 120 ディスクアダプタ
- 130 キャッシュメモリ
- 140 管理メモリ
- 150 記憶装置
- 160 コモンパス
- 170 ディスクアダプタと記憶装置間の接続線
- 180 ホストコンピュータ



【書類名】図面【図1】

図 1 ホストコンピュータ 副記憶システムB 180 100B 200 CHA CHA CACHE SM<u>110</u> <u>110</u> <u>130</u> 140 160 DKA DKA DKA DKA 120 120 120 120 HDDHDD 170 170 150 **150** HDD HDD 150 150 HDD HDD 正記憶システムA <u>150</u>

100A

[図2]

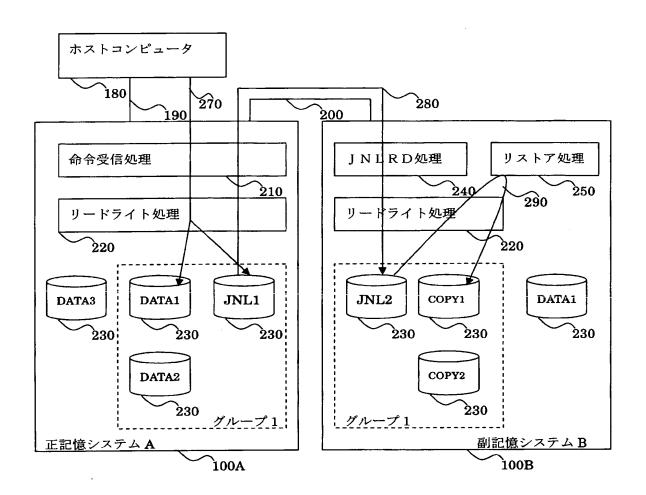
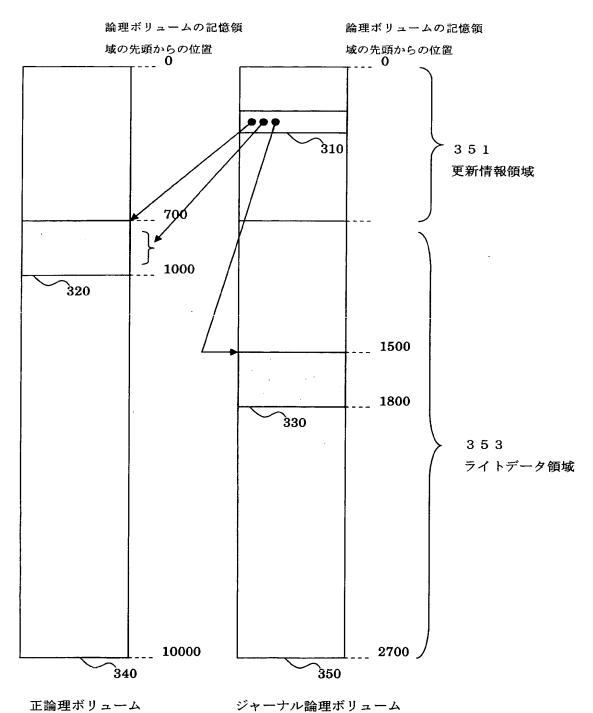




図 3



【図4】

⊠ 4

| 設定項目 | 設定値例 |
|-------------------------|---------------------------|
| ライト命令を受信した時刻 | 1999/3/17 22:20:10 |
| グループ番号 | |
| 更新番号 | 4 |
| ライト命令の論理アドレス | 論理ボリューム番号:1 |
| | 論理ボリュームの記憶領域の先頭からの位置:700 |
| ライトデータのデータサイズ | 300 |
| ライトデータを格納したジャーナル論理ボリューム | 論理ボリューム番号:4 |
| の論理アドレス | 論理ボリュームの記憶領域の先頭からの位置:1500 |

310 更新情報

【図5】

| | | | | | | | | | | | 400 ボリューム情報 |
|--------|--------|----------------------------|--------|-----|-------|-------|-------|-------|--------------|-------|----------------|
| | K | | 先頭から位 | | 0 | က | 6 | 0 | 6 | 12 | 400 ₩ U |
| | 物理アドレス | | 記憶装 | 置番号 | -1 | | - | 2 | 2 | 2 | |
| | ペア番号 | | | | 1 | 2 | 0 | 0 | 0 | 0 | |
| ∑ ∑ | 容量 | (GB) | | | 3 | 9 | 9 | 6 | 3 | 9 | |
| | フォー | マップ・インド | 7+01 | | OPEN3 | OPEN6 | OPEN6 | OPEN9 | OPEN3 | OPEN6 | |
| | ボリューム | 大丽 | | | 王 | 田 | 未使用 | 正常 | 正常 | 未使用 | |
| | 警進 | ボ フェ ト 本 中 本 中 | l E | | 1 | 2 | 3 | 4 | 5 | 9 | |

出証特2004-3002965

【図6】

| | _ | , | , | | _ | , | , |
|------------|----------------------|----|----|-----|-----|--------------|---|
| | し プロ ア スファ | 0 | 0 | 0 | 0 | 0 | |
| | グループ番号 | 1 | 1 | 0 | 0 | 0 | との ペア声描 |
| | 副 当 リューム 番号 | | က | 0 | 0 | 0 | \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ |
| | 副記憶システム番号 | 2 | 2 | 0 | 0 | 0 | |
| ⊠ 6 | 正計選用ボリューム | 1 | 2 | 0 | 0 | 0 | |
| | 正記憶システム番号 | 1 | - | 0 | 0 | 0 | |
| | ペア状態 | 正常 | 正常 | 未使用 | 未使用 | 未使用 | |
| | ペア番号 | 1 | 2 | 3 | 4 | 5 | |

出証特2004-3002965

【図7】

<u>⊠</u>

| | | | 青穀 |
|--------------------|-----|-----|------------|
| 更新番号 | Ŧ | 0 | 600 グループ情報 |
| ジャーナル 論理ボリューム番号 | 4 | 0 | |
| ペア集合 | 1,2 | 0 | |
| グループ状態 | 正常 | 未使用 | |
| み たープ 番号 | 1 | 2 | |

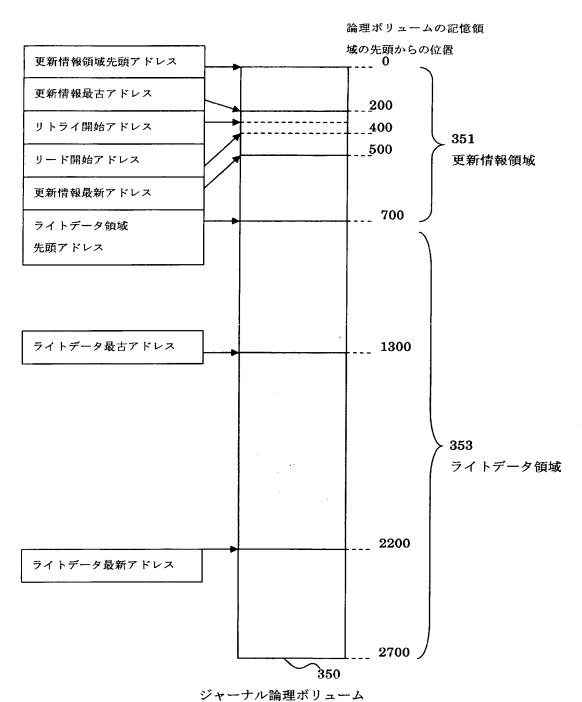
【図8】

| | 8 | | |
|----------------|---------|---------------|-------------|
| | 副記憶システム | 満埋アドレス | |
| | 番号 | 論理ボリューム番号 | 熱理ボリュームの記憶領 |
| | | | 域の先頭からの位置 |
| 更新情報領域先頭アドレス | 2 | 4 | 0 |
| ライトデータ領域先頭アドレス | 2 | 4 | 700 |
| 更新情報最新アドレス | 2 | 4 | 500 |
| 更新情報最古アドレス | 2 | 4 | 200 |
| ライトデータ最新アドレス | 2 | 4 | 2200 |
| ライトデータ最古アドレス | 2 | 4 | 1300 |
| リード開始アドレス | 2 | 4 | 400 |
| リトライ開始アドレス | 2 | 4 | 300 |
| | | | 700 ポインタ情報 |

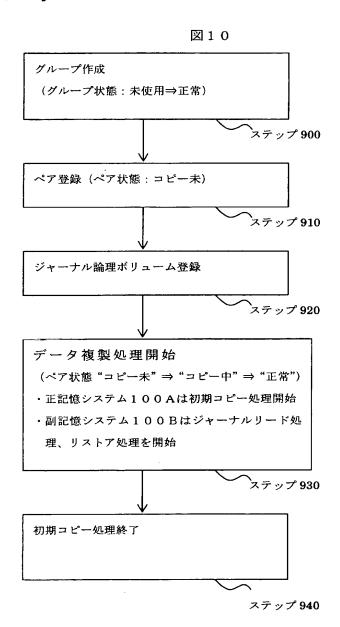
出証特2004-3002965

【図9】

図 9

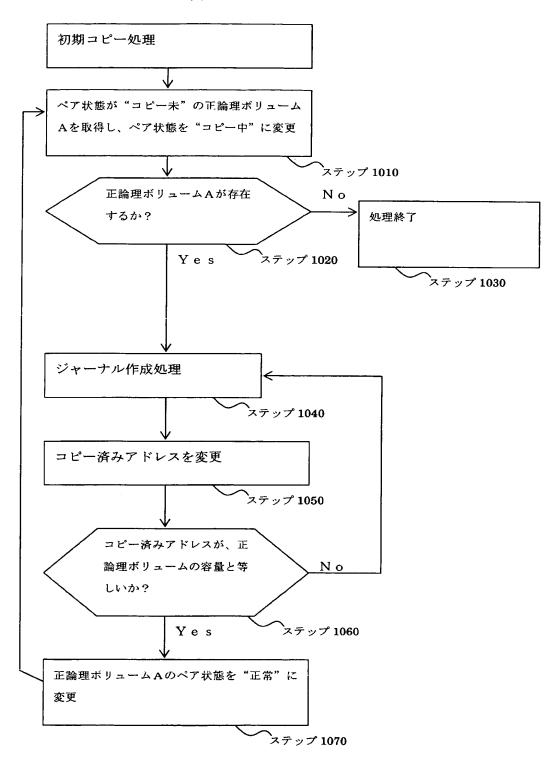


【図10】

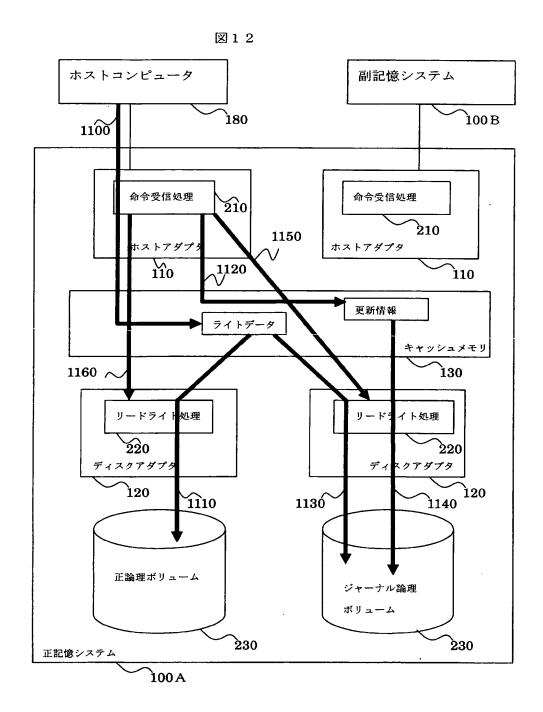


【図11】

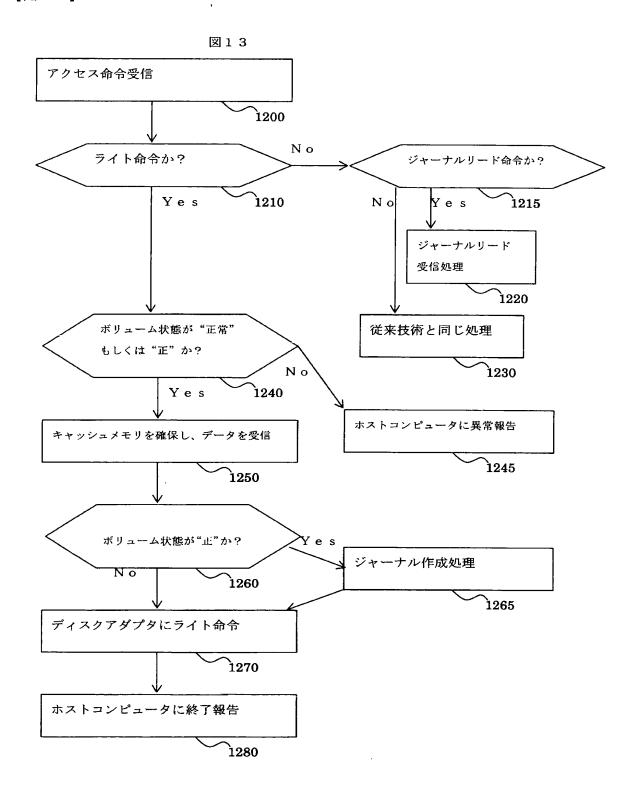
図11



【図12】



【図13】



【図14】

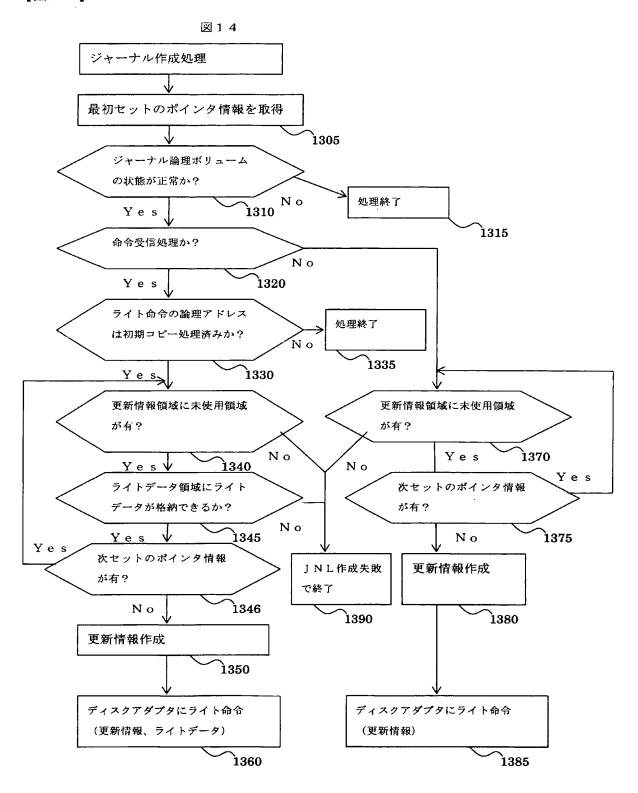
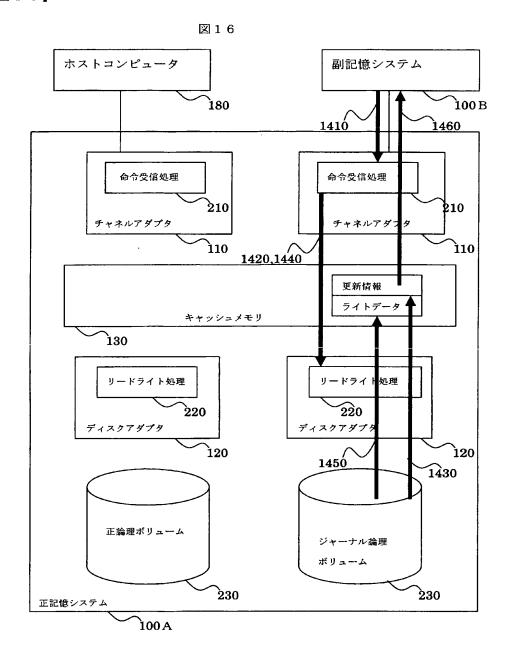


図15

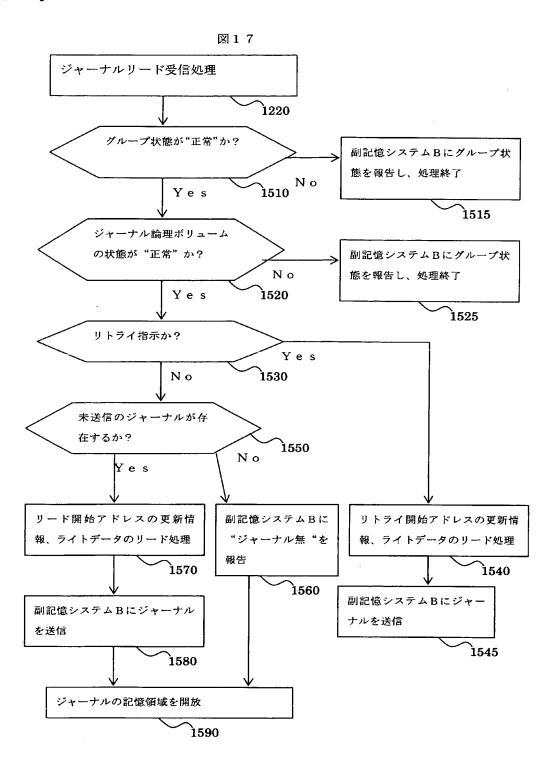
| ライト命令を受信した時刻1999/3/17 22:22:10グループ番号1更新番号4ライト命令の論理アドレス論理ボリューム番号: 1論理ボリューム番号: 1論理ボリューム番号: 1 | 22:10 |
|--|----------------------------|
| ループ番号 新番号 イト命令の論理アドレス | |
| K7 | |
| レメ | |
| | 人番号:1 |
| | 論理ボリュームの記憶領域の先頭からの位置:800 |
| ライトデータのデータサイズ 100 | |
| ライトデータを格納したジャーナル論理ボリューム 論理ボリューム番号:4 | 人番号:4 |
| の論理アドレス 論理ボリュームの | 論理ボリュームの記憶領域の先頭からの位置: 2200 |

10 更新情報

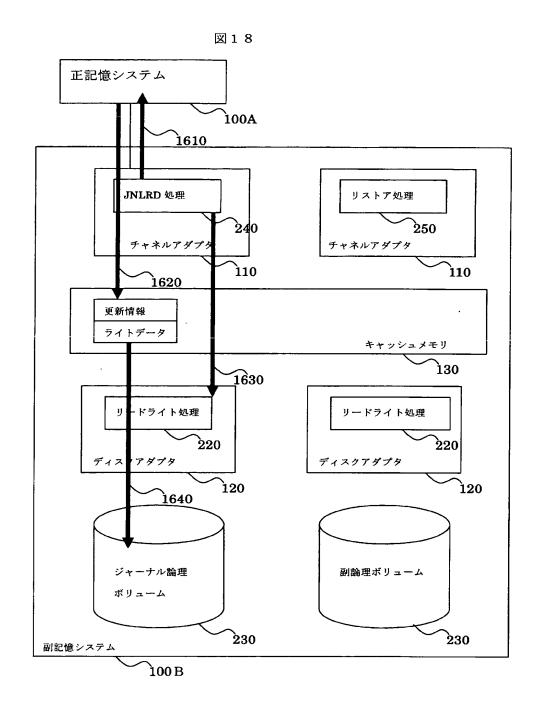
【図16】



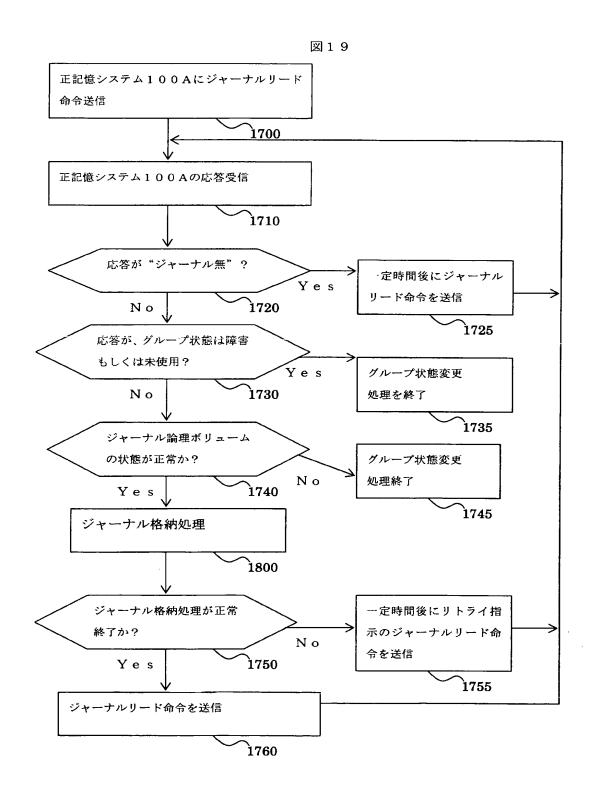
【図17】



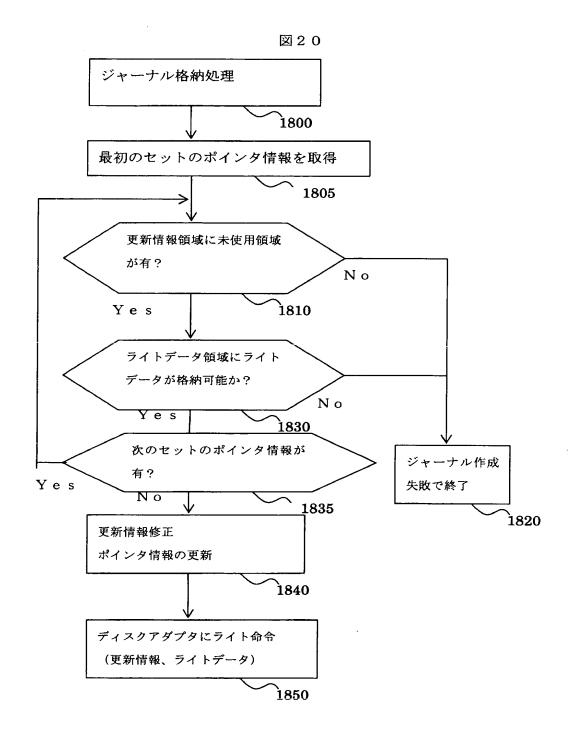
【図18】



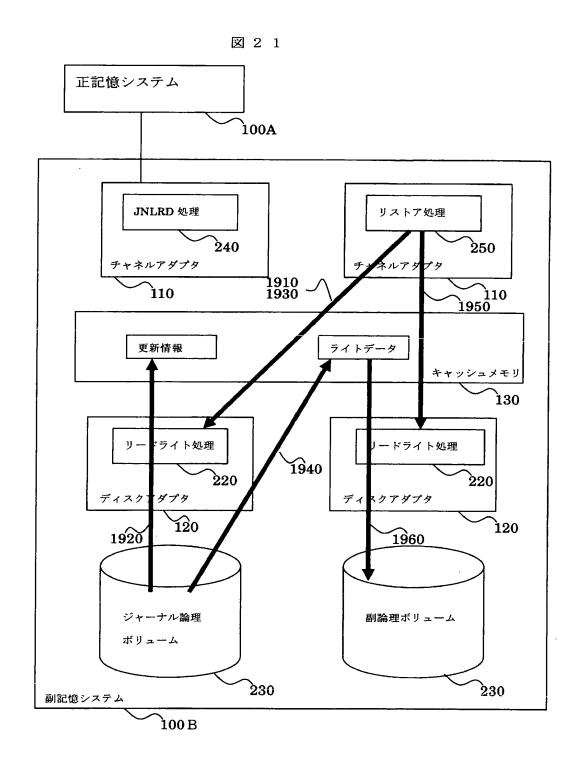
【図19】



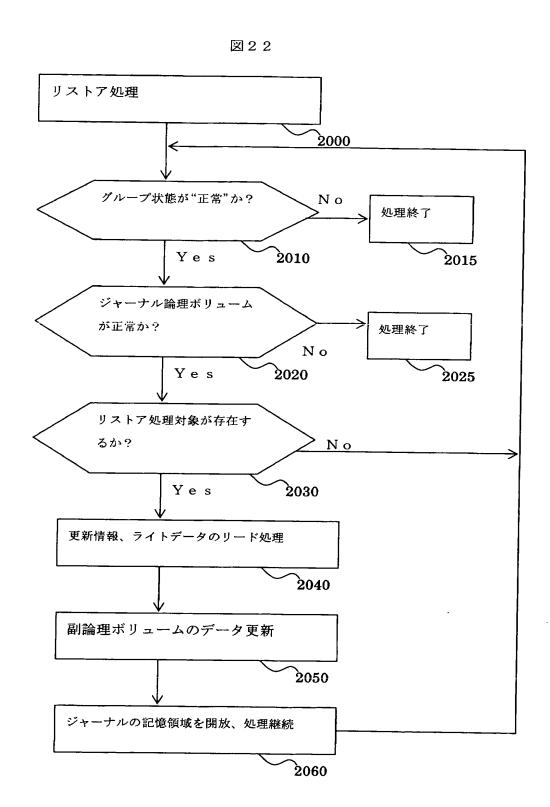
【図20】



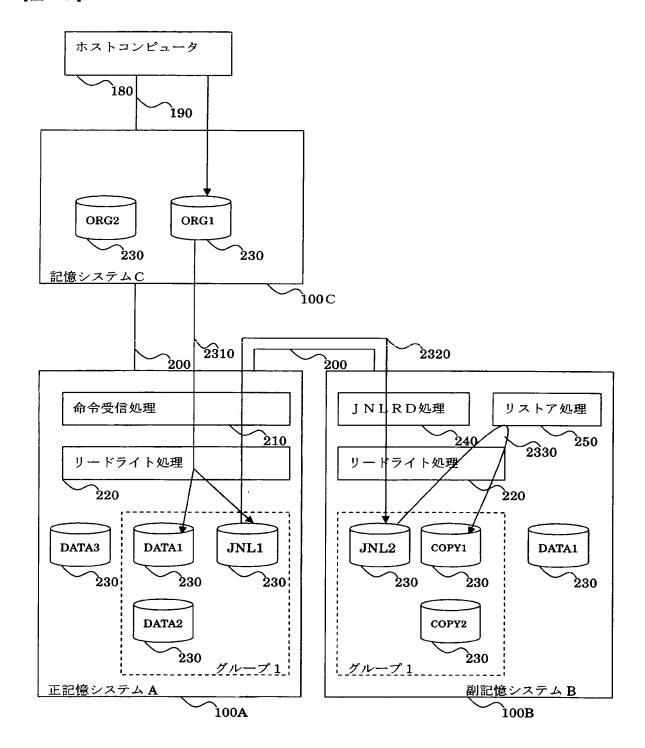
【図21】



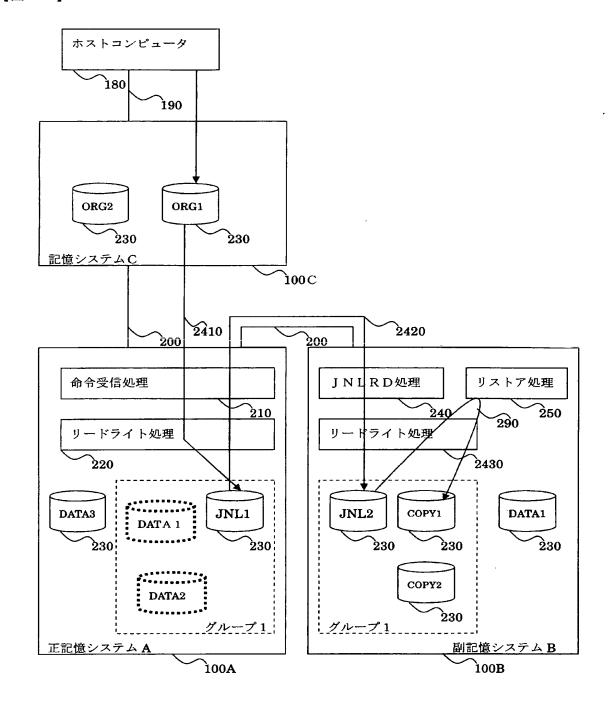
【図22】



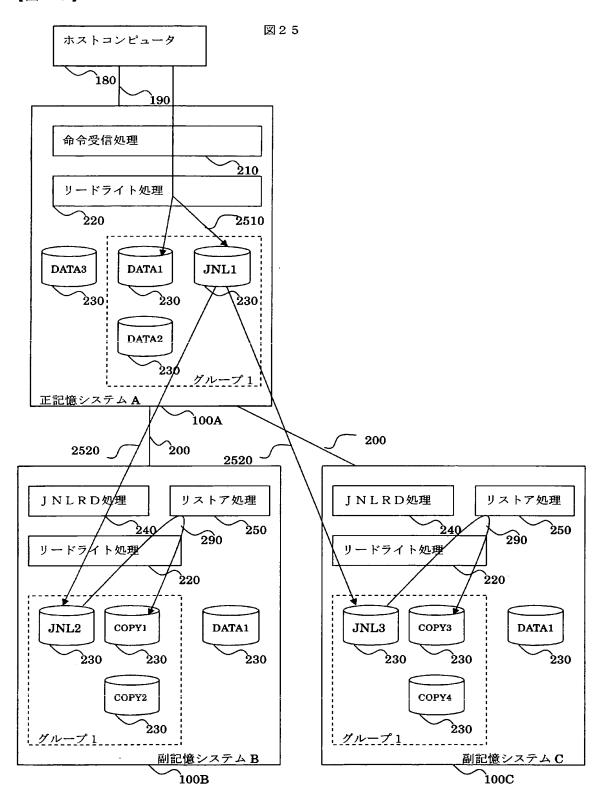
【図23】



【図24】



【図25】



【図26】

図26

| | 副記憶シス | 論理アドレス | |
|----------------|-------|-----------|--------------------------|
| | テム番号 | 論理ボリューム番号 | 論理ボリュームの記憶領 域の先頭からの位置 |
| 更新情報領域先頭アドレス | 2 | 4 | 0 |
| ライトデータ領域先頭アドレス | 2 | 4 | 700 |
| 更新情報最新アドレス | 2 | 4 | 500 |
| 更新情報最古アドレス | 2 | 4 | 200 |
| ライトデータ最新アドレス | 2 | 4 | 2200 |
| ライトデータ最古アドレス | 2 | 4 | 1300 |
| リード開始アドレス | 2 | 4 | 400 |
| リトライ開始アドレス | 2 | 4 | 300 |

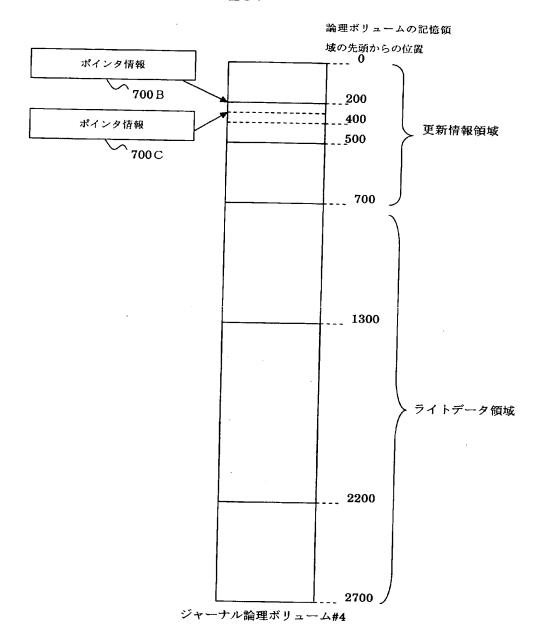
↑ 700B ポインタ情報

| | 副記憶シス | 論理アドレス | |
|----------------|-------|-----------|--------------------------|
| | テム番号 | 論理ボリューム番号 | 論理ボリュームの記憶領 域の先頭からの位置 |
| 更新情報領域先頭アドレス | 3 | 4 | 0 |
| ライトデータ領域先頭アドレス | 3 | 4 | 700 |
| 更新情報最新アドレス | 3 | 4 | 500 |
| 更新情報最古アドレス | 3 | 4 | 300 |
| ライトデータ最新アドレス | 3 | 4 | 2200 |
| ライトデータ最古アドレス | 3 | 4 | 1500 |
| リード開始アドレス | 3 | 4 | 500 |
| リトライ開始アドレス | 3 | 4 | 400 |

[↑]700C ポインタ情報

【図27】

図27



【書類名】要約書

【要約】

【課題】 記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ 転送又はデータの複製を行う。

【解決手段】 上位装置 180 に接続された正記憶システム 100 Aに、 26 以上の副記憶システム 100 B、 100 Cが接続される。副記憶システム 100 B、 100 Cは、それぞれ、独自のタイミングで正記憶システム 100 A からデータ更新のジャーナルを読み所定の論理ボリューム J N L 2、 J N L 3 に保存し、そして、独自のタイミングで論理ボリューム J N L 2、 J N L 3 内のジャーナルに基づき、正記憶システム 100 A 内のデータの複製を生成して、副論理ボリューム C O P Y 1、 C O P Y 3 に保存する。正記憶システム 100 A は、副記憶システム 100 B、 100 C の双方がジャーナルを読んでリストアするまで、そのジャーナルを保持する。ジャーナルリードのタイミングはジャーナル量、処理負担などに応じて制御される。

【選択図】図25

認定・付加情報

特許出願の番号

特願2003-416414

受付番号

5 0 3 0 2 0 6 0 6 5 0

書類名

特許願

担当官

第七担当上席

0096

作成日

平成15年12月16日

<認定情報・付加情報>

【提出日】

平成15年12月15日



特願2003-416414

出願人履歴情報

識別番号

[000005108]

1. 変更年月日 [変更理由] 住 所

1990年 8月31日

新規登録

東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所